The Nucleosome Remodelling and Deacetylation complex suppresses transcriptional noise during lineage commitment

Thomas Burgold[1,4], Michael Barber[1], Susan Kloet[2, 5], Julie Cramard[1], Sarah Gharbi[1], Robin Floyd[1], Masaki Kinoshita[1], Meryem Ralser[1], Michiel Vermeulen[2], Nicola Reynolds[1], Sabine Dietmann[1] and Brian Hendrich[1, 3]

1. Wellcome– MRC Stem Cell Institute, University of Cambridge, Cambridge CB2 1QR United Kingdom
2. Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University, 6525 GA Nijmegen, The Netherlands
3. Department of Biochemistry, University of Cambridge, Cambridge CB2 1QR United Kingdom
4. Present address: Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom
5. Present address: Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands

Corresponding Author:

Brian Hendrich, Brian.Hendrich@cscr.cam.ac.uk, +44 (0)1223 760205, @BDH_Lab

Running Title: NuRD suppresses transcriptional noise

Abstract

Multiprotein chromatin remodelling complexes show remarkable conservation of function amongst metazoans, even though components present in invertebrates are often found as multiple paralogous proteins in vertebrate complexes. In some cases these paralogues specify distinct biochemical and/or functional activities in vertebrate cells. Here we set out to define the biochemical and functional diversity encoded by one such group of proteins within the mammalian Nucleosome Remodelling and Deacetylation (NuRD) complex: Mta1, Mta2 and Mta3. We find that, in contrast to what has been described in somatic cells, MTA proteins are not mutually exclusive within ES cell NuRD and, despite subtle differences in chromatin binding and biochemical interactions, serve largely redundant functions. ES cells lacking all three MTA proteins represent a complete NuRD null and are viable, allowing us to identify a previously unreported function for NuRD in reducing transcriptional noise, which is essential for maintaining a proper differentiation trajectory during early stages of lineage commitment.

Key Words: Chromatin, ES Cell, Lineage Commitment, NuRD, Transcription

Introduction

Mammalian cells contain a number of proteins capable of using ATP hydrolysis to shift nucleosomes relative to the DNA sequence, thereby facilitating chromatin remodelling. In mammals, these ATP-dependent chromatin remodelling proteins usually exist within multiprotein complexes and play essential roles in the control of gene expression, DNA replication and repair (Hargreaves & Crabtree, 2011, Hota & Bruneau, 2016, Narlikar et al., 2013).

NuRD (Nucleosome Remodelling and Deacetylation) is one such multiprotein complex which is unique in that it contains both chromatin remodelling and protein deacetylase activity. NuRD is highly conserved amongst metazoans and has been shown to play important roles in cell fate decisions in a wide array of systems (Denslow & Wade, 2007, Signolet & Hendrich, 2015). For example, in embryonic stem (ES) cells NuRD controls nucleosome positioning at regulatory sequences to finely tune gene expression (Bornelov et al., 2018, Reynolds et al., 2012) and in somatic lineages NuRD activity has been shown to prevent inappropriate expression of lineage-specific genes to ensure fidelity of somatic lineage decisions (Denner & Rauchman, 2013, Gomez-Del Arco et al., 2016, Knock et al., 2015, Loughran et al., 2017). It was recently demonstrated that this is achieved in both ES cells and B-cell progenitors by restricting access of transcription factors to regulatory sequences (Bornelov et al., 2018, Liang et al., 2017, Loughran et al., 2017). Additionally, aberrations in expression levels of NuRD component proteins are increasingly being linked to cancer progression (Lai & Wade, 2011, Mohd-Sarip et al., 2017).

NuRD is comprised of two enzymatically and biochemically distinct subcomplexes: a chromatin remodelling and a deacetylase subcomplex. The chromatin remodelling subcomplex contains a nucleosome remodelling ATPase protein (Chd3/4/5) along with one

of the zinc finger proteins Gatad2a/b and the Doc1/Cdk2ap1 protein, while the deacetylase subcomplex contains class I histone deacetylase proteins Hdac1/2, the histone chaperones Rbbp4/7, the Metastasis Tumour Antigen family of proteins, Mta1, Mta2 and Mta3 and, in pluripotent cells, the zinc finger proteins Sall1/4 (Allen et al., 2013, Bode et al., 2016, Kloet et al., 2015, Lauberth & Rauchman, 2006, Low et al., 2016, Miller et al., 2016, Spruijt et al., 2016, Zhang et al., 2016). These two subcomplexes are bridged by Mbd2/3, creating intact NuRD (Fig. 1A). While HDAC and RBBP proteins are also associated with other chromatin modifying complexes, the MBD, GATAD2 and MTA proteins are obligate NuRD components. Functional and genetic data indicate that the CHD4-containing remodelling subunit is capable of functioning independently of intact NuRD (O'Shaughnessy & Hendrich, 2013, O'Shaughnessy-Kirwan et al., 2015, Ostapcuk et al., 2018). The deacetylase subcomplex has been shown biochemically to exist outside of intact NuRD, though whether this subcomplex has any specific function is not clear (Link et al., 2018, Zhang et al., 2018).

Changes in subunit composition in large multiprotein, chromatin modifying complexes such as PRC1 and BAF have been shown to correlate with distinct changes in function to sites of action in the chromatin in a cell-type specific manner (Ho & Crabtree, 2010, Morey et al., 2012). The NuRD complex might therefore be expected to show similar diversity in both composition and function and in fact, diversification of NuRD function has been described through differential incorporation of various isoforms of NuRD component proteins (Bowen et al., 2004). For example, Mbd2 and Mbd3 are mutually exclusive within NuRD (Le Guezennec et al., 2006). While Mbd2 is not required for mammalian development, Mbd3 is essential at early postimplantation stages in mice (Hendrich et al., 2001). Mbd2/NuRD is a methyl-CpG binding co-repressor complex which is dispensable for early development but Mbd3/NuRD, a transcriptional modulator found at sites of active

transcription, has been shown to play important roles in regulation of cell fate decisions in multiple developmental systems (Feng & Zhang, 2001, Gunther et al., 2013, Menafra et al., 2014, Reynolds et al., 2012, Reynolds et al., 2013, Shimbo et al., 2013). NuRD complexes containing either CHD3, CHD4 or CHD5 play distinct roles during cortical development (Nitarska et al., 2016).

Further functional and biochemical diversification occurs through alternate use of the three MTA proteins within NuRD. MTA proteins function as a scaffold around which the deacetylase subcomplex is formed, comprising a 2:2:4 stoichiometry of MTAs:HDACs:RBBPs (Millard et al., 2016, Millard et al., 2013, Smits et al., 2013, Zhang et al., 2016). While invertebrates predominantly have a single MTA protein orthologue, most vertebrates have three. The three mammalian MTA proteins are highly conserved, differing from each other predominantly at their C-termini. The MTA1 protein was originally identified because of its elevated expression in metastatic cell lines (Toh et al., 1994), and subsequently all three MTA proteins have been shown to be up-regulated in a range of different cancer types (Covington & Fuqua, 2014, Ma et al., 2016, Sen et al., 2014). The MTA1 and 3 proteins were shown to form distinct NuRD complexes in breast cancer cells and in B-cells and were recruited by different transcription factors to regulate gene expression (Fujita et al., 2004, Fujita et al., 2003, Mazumdar et al., 2001, Si et al., 2015). These studies did not detect biochemical interactions between MTA3 and the other MTA proteins, leading to the conclusion that MTA proteins are mutually exclusive within NuRD. In contrast, Mta1 was shown to interact with Mta2 in MEL cells, possibly indicating that mutual exclusivity may be cell type-specific (Hong et al., 2005). While all three *Mta* genes are expressed in ES cells, detailed biochemical analysis of interactions of MTA proteins with one another or with the various NuRD components in ES cells has not previously been described.

Functional evidence does not support a strict lack of redundancy amongst MTA proteins during mammalian development. While zygotic deletion of *Chd4* or *Mbd3* results in pre- or peri-implantation developmental failure respectively (Kaji et al., 2007, O'Shaughnessy-Kirwan et al., 2015), mice deficient in any one of the three MTA proteins show minimal phenotypes. Mice lacking either *Mta1* or *Mta3* are viable and fertile (Manavathi et al., 2007)(Mouse Genome Informatics), while mice lacking *Mta2* show incompletely penetrant embryonic lethality and immune system defects (Lu et al., 2008). In the current study we took a systematic approach to dissecting MTA protein biochemical and functional diversity. We find that, in contrast to what has been described in somatic cells, MTA proteins are not mutually exclusive within NuRD in ES cells and serve largely redundant functions. Furthermore, ES cells lacking all three MTA proteins are viable and represent a complete NuRD null, allowing us to identify a previously undetected function for NuRD in early stages of lineage commitment.

Results

MTA proteins are not mutually exclusive within the NuRD complex in ES cells

The absence of a detected interaction between the MTA2 and MTA3 proteins in human cells (Fujita et al., 2003, Si et al., 2015), and the observation that different MTA proteins can show different protein-protein interactions in B-cells (Fujita et al., 2004) has led to the conclusion that the MTA proteins are mutually exclusive within NuRD, and could hence confer functional diversity to the NuRD complex (Lai & Wade, 2011). To investigate the biochemical specificity of the MTA proteins in an unbiased manner, we used gene targeting of endogenous loci to produce three different mouse ES cell lines in which an epitope tag was fused to the C-terminus of each MTA protein (Fig. 1A). Although MTA genes

show different expression patterns in preimplantation mouse development, all three are expressed in peri implantation and early post implantation epiblast, the tissue most similar to the naïve ES cell state (Fig. EV1). We therefore considered ES cells to be a good system in which to investigate the function of MTA proteins.

Each tagged protein was expressed at levels comparable to those of wild type proteins and was found to interact with other NuRD component proteins by immunoprecipitation (Fig. 1B). Each MTA protein was also able to immunoprecipitate the other MTA proteins in addition to unmodified forms of itself (Fig. 1B). Mta3 was barely detectable in Mta1-3xFLAG immunoprecipitates by IP-western blot (though robustly found by mass spectrometry, see below), but Mta1 was robustly detected associating with Mta3-3xFLAG (Fig. 1B). This could indicate that any Mta1-Mta3 containing NuRD complexes represent a relatively small proportion of nuclear Mta1, but a relatively large proportion of nuclear Mta3. Individual NuRD complexes contain two MTA proteins (Millard et al., 2013, Smits et al., 2013, Zhang et al., 2016), so the identification of an interaction between MTA proteins means that NuRD complexes in ES cells could contain either homodimers or heterodimers consisting of any combination of the three MTA proteins.

To investigate the potential biochemical diversification conferred by different MTA proteins in pluripotent cells we used our tagged cell lines to identify proteins interacting with each of the MTA proteins by label-free quantitative mass spectrometry. Each protein robustly co-purified with all known NuRD component proteins, including each of the MTA proteins, confirming that MTA proteins are not mutually exclusive within NuRD in ES cells (Fig. 1C). In addition to NuRD components, each of the MTA proteins co-purified with Wdr5 as well as a number of zinc finger proteins, most of which had previously been identified as

NuRD-interacting proteins (Bode et al., 2016, Ee et al., 2017, Matsuura et al., 2017, Spruijt et al., 2016).

As NuRD is assembled from a deacetylase subcomplex and a remodeller subcomplex joined through the Mbd3 protein (Fig. 1A), loss of Mbd3 is expected to result in dissociation of these two subcomplexes. Endogenous tagging for each Mta protein was performed in an ES cell line harbouring a floxed *Mbd3* allele, and IP/Mass spectrometry was repeated in ES cells after *Mbd3* deletion in order to enrich for interactions specific for the deacetylase subcomplex. The majority of interactions with non-NuRD components was lost after *Mbd3* deletion, indicating that most of these proteins do not directly associate with either the MTA proteins or with the deacetylase subcomplex (Fig. 1C). Exceptions to this were Zfp296, which was identified as interacting with all three MTA proteins in an *Mbd3*-independent manner, and Pwwp2a, which co-purified with Mta1 in both wild type and Mbd3-null cells, consistent with recent reports (Link et al., 2018, Zhang et al., 2018). Notably, Zfp219 and Zfp512b both showed *Mbd3*-independent interactions specifically with Mta2 (Fig. 1C).

By quantitating the abundance of peptides sequenced in each experiment we found that the interactomes for both Mta1 and Mta2 showed a depletion of peptides associated with the remodeller subcomplex (i.e. Chd4, Gatad2a/b and Cdk2ap1) in *Mbd3*-null cells, whereas proteins associated with the histone deacetylase subcomplex (Mta proteins, Hdac1/2, Sall proteins and Rbbp4/7) remained present at similar levels (Fig 1D). In contrast the Mta3 interactome showed an increased interaction with Mbd2 and no relative loss of either sub-complex in the absence of Mbd3. Thus both Mta1 and Mta2 can form part of a stable histone deacetylase-containing subcomplex in the absence of intact NuRD, but Mta3 is preferentially found in intact NuRD complexes. All three MTA proteins associated with Mbd2 in the absence of Mbd3, indicating that they can all can contribute to Mbd2/NuRD.

Together these data show that the MTA proteins are found exclusively within the NuRD complex in ES cells.

MTA proteins show subtle differences in genome-wide binding patterns

To test whether MTA proteins confer differential chromatin binding to NuRD complexes we subjected our ES cell lines expressing epitope-tagged MTA proteins to chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq). The binding profiles of the three proteins were largely, but not completely overlapping (Fig. 2A, B). Mta3 binding was almost entirely associated with Mta1 and/or Mta2-binding, while 31% of Mta1 peaks and 22% of Mta2 peaks were not associated with any other MTA protein (Fig. 2A,B). The overall efficiency of Mta3-3xFLAG ChIP-seq was lower than those for Mta1-3xFLAG or Mta2-GFP, possibly due to the lower levels of Mta3 present in ES cells, although replicates were highly correlated (Appendix Figure S1). While Mta3-bound sites can therefore be identified with high confidence, it remains formally possible that “Mta1-only” or “Mta2-only” sites also contain Mta3, albeit at levels beneath the detection threshold of our ChiP-seq. The vast majority of MTA peaks were also bound by Chd4, indicating that they represent NuRD-bound regions (Fig. 2C, D; Fig. EV2A, B).

Sites found associated with all three MTA proteins were highly enriched for Chd4 and Mbd3 binding, consistent with these being core NuRD binding regions (Fig. 2D). These sites were also enriched for marks of active promoters (H3K4Me3, H3K27Ac) and active enhancers (H3K4Me1, H3K27Ac, P300; Fig. 2D, E), both of which are hallmarks of NuRD-associated regions (Bornelov et al., 2018, Miller et al., 2016). The same was true for sites bound by any two of the three MTA proteins (Fig. EV2B). The majority of sites occupied by one MTA protein, but not the other two, were also occupied by Chd4 but to a lesser extent

than is seen at core NuRD sites (Fig. 2D, E). Sites associated with all three MTA proteins, presumably representing core NuRD-bound sites, were located both at transcription start sites and at distal locations, while sites bound by only one MTA protein were predominantly located distal to transcription start sites (Fig EV2C). Whereas Mta2-only and Mta3-only sites showed some enrichment for both H3K4Me1 and H3K4Me3, Mta1-only sites showed no enrichment for H3K4Me3 (Fig. 2D, E). In all cases MTA-only sites lacking Chd4, which could represent sites of binding by the histone deacetylase subcomplex only, showed characteristics of inactive enhancers, in that they were moderately enriched for H3K4Me1 and P300, but not for H3K4Me3, H3K27Ac or H3K36Me3 (Fig EV2D). None of the MTA-bound sequences showed any enrichment for methylated DNA (Fig. EV2E). Genes associated with binding by only one or two MTA proteins, with or without Chd4 are associated with a similar distribution of GO terms (Fig EV2F) which is not consistent with the idea that the MTA proteins are directing NuRD activity to specific gene subsets.

Mta1/Mta2/Mta3 triple knockout is a total NuRD null

To investigate the degree of genetic redundancy amongst the different MTA genes, we obtained gene trap alleles for each gene from the European Conditional Mouse Mutagenesis Programme (Skarnes et al., 2011) as ES cell lines (*Mta1* and *Mta2*) or as embryos (*Mta3*) (Fig. EV3A-C). ES cell lines were used for morula aggregation to create chimaeric mice which were subsequently outcrossed to establish a mouse line. Conditional deletion alleles were generated for each line, which was subsequently bred with females expressing Sox2-Cre (Hayashi et al., 2002) resulting in deletion of floxed exons and the absence of any detectable protein production from each allele (Fig. EV3D; see Methods for details). As has been reported previously, *Mta1*$^{-/-}$ and *Mta3*$^{-/-}$ mice were viable, and Mta2$^{-/-}$

mice showed incompletely penetrant embryonic lethality (Lu et al., 2008, Manavathi et al., 2007).

Mta1, *Mta2* or *Mta3*-null ES cell lines derived from mice were morphologically indistinguishable from wild type ES cells, as were ES cell lines deficient for combinations of pairs of MTA genes (Fig. EV3E). While some differences in gene expression could be identified between single mutants and their respective parent lines (Fig EV3F, G), these changes were not sufficient to make them distinct from wild type ES cells when compared to *Mbd3*-mutant ES cells, 3.5 dpc ICM and 4.5 dpc epiblast (Fig EV3H). We therefore concluded that any gene expression changes identified in single mutant ES cells were unlikely to be of consequence during early development, and that MTA proteins show functional redundancy during early mammalian development.

Mta1$^{-/-}$*Mta2*$^{-/-}$*Mta3*$^{Flox/Flox}$ ES cells (*Mta12Δ*) were subsequently created and expanded in culture. After transfection with a Cre expression construct to induce deletion of both *Mta3* alleles we recovered ES cells lacking all three MTA proteins (see Methods for details; Fig. 3A). These *Mta1*$^{-/-}$*Mta2*$^{-/-}$*Mta3*$^{-/-}$ ES cells (subsequently referred to as *Mta123Δ*) appeared morphologically normal in standard, 2i + LIF culture (Fig. 3B), but showed a considerable degree of spontaneous differentiation in serum/LIF culture. MTA proteins are therefore dispensable for ES cell viability.

Structurally, MTA proteins bridge an interaction between the deacetylase subcomplex with Mbd3 and the remodelling subcomplex (Fig. 1A) so we predicted that loss of all three MTA proteins would prevent NuRD formation. Consistent with this prediction we could detect no interactions between Chd4 and components of the deacetylase subcomplex (Hdac2, Rbbp4) in *Mta123Δ* ES cell nuclear extract by immunoprecipitation of endogenous proteins (Fig. 3C). Surprisingly, despite being transcribed at normal levels, both Gatad2b and

Mbd3 proteins were present at reduced levels in *Mta12*Δ cells and barely detectable in *Mta123Δ* cells (Fig. 3C; Fig. EV4A), indicating that the MTAs are important for the stability of both of these proteins. To investigate this further we monitored loss of protein stability in $Mta1^{\Delta/\Delta}Mta2^{\Delta/\Delta}Mta3^{Flox/Flox}$ ES cells after deletion using a tamoxifen-inducible Cre (Fig. 3D). Loss of Mbd2, Mbd3, Gatad2a and Gatad2b protein stability were all coincident with loss of Mta3 protein, although Mbd3 and Gatad2b were already at reduced levels in the starting cell line ($Mta12\Delta Mta3^{Flox/Flox}$) (Fig. 3D). Gatad2b and Mta3 also show reduced protein stability in *Mbd3*Δ ES cells, indicating that a particular inter-dependency exists between these three proteins for stability (Fig EV4B). Furthermore, introduction of either Mta1, Mta2 or Mta3 into *Mta123Δ* ES cells at levels comparable to wild type expression resulted in restoration of Mbd3 protein levels, demonstrating that contact with at least one of the MTA proteins is sufficient for Mbd3 protein stability (Fig. 3E). In contrast to *Mbd3*-null ES cells which display a significant depletion, but not a complete loss of NuRD due to partial compensation by Mbd2 (Kaji et al., 2006), *Mta123Δ* ES cells were completely devoid of any detectable intact NuRD (Fig. 3C, F). The *Mta123Δ* ES cells are thus a total NuRD null ES cell line, which allowed us to examine, for the first time, the consequences of a complete loss of the NuRD complex in a viable mammalian cell system.

NuRD suppresses transcriptional noise

Twice as many genes showed an increase rather than a decrease in expression levels in *Mta123Δ* ES cells compared to control ES cells by RNA-seq (Fig. 4A). The ratio of increased to decreased gene expression in *Mta123Δ* cells was very similar for genes bound by all three Mta proteins, Mta1+2, Mta1 only or Mta2 only (Fig. 4B). NuRD's global impact on

transcription is therefore not grossly altered by the inclusion or absence of individual MTA subunits.

Comparing global gene expression profiles using principal component analysis (PCA) showed that *Mta123Δ* and *Mbd3Δ* ES cells in self-renewal conditions were more similar to each other than either was to wild type cells (Fig. 4C). *Mta123Δ* ES cells were most distinct from wild type cells, consistent with them representing a more complete NuRD knockout (Fig. 4C). Consistent with this, genes driving the NuRD-specific separation (Principal Component 2; PC2) were generally misexpressed to a greater degree in *Mta123Δ* ES cells than in *Mbd3Δ* ES cells (Fig EV4C). Genes misexpressed in either *Mbd3Δ* or *Mta123Δ* ES cells were most significantly associated with the same top three GO terms, but the significance was greater in *Mta123Δ* ES cells as these cells misexpressed more genes in each category and generally to a greater extent than did the *Mbd3Δ* cells (Fig. 4D, E, EV4C).

Genes misexpressed in *Mta123*Δ ES cells were predominantly those normally expressed at lower levels (Fig. EV4D). A very similar pattern was seen if we considered genes misexpressed in *Mbd3*Δ ES cells, with the major difference being that more genes were misexpressed in the *Mta123*Δ cells. The predominant biochemical difference between the *Mbd3*Δ cells and *Mta123*Δ cells was the lack of Mbd2/NuRD and of the MTA/HDAC subcomplex in the triple mutants, so we next asked whether we could detect evidence for specific functions for either of these complexes in ES cells. We identified genes near core NuRD peaks (N=15210), near Mta-only peaks (regardless of the presence of a NuRD peak; N=10280), genes associated with NuRD components Chd4 and MTAs, but not Mbd3 (presumptive Mbd2/NuRD-only genes; N=349), and genes containing MTA-only peaks, but no Chd4 peak (MTA-only genes; N=151) (Fig. 4F) and divided them into expression quartiles. While there are very few genes not associated with a NuRD peak, those few presumptive

Mbd2/NuRD-only genes or MTA-only genes showed a similar misexpression pattern to those seen for core NuRD genes, in that the majority of misexpressed genes was in the second lowest expression quartile (Fig. 4F). We therefore find no evidence for a specific gene regulatory function associated with an MTA/HDAC subcomplex. This is consistent with a recent report in which loss of Pwwp2a/b resulted in only very slight alterations in nascent RNA production amongst a small subset of genes (Zhang et al., 2018). Nevertheless, this analysis further highlights the importance for NuRD in suppressing expression of lowly expressed genes. We therefore conclude that, in addition to fine-tuning gene expression as part of the Mbd3/NuRD complex, MTA proteins are also important for preventing inappropriate activation of a variety of different genes, or in preventing transcriptional noise in ES cells.

The NuRD complex safeguards cellular identity during differentiation

We next asked what impact complete loss NuRD activity had upon the differentiation capacity of *Mta123Δ* ES cells. Upon removal of the two inhibitors and LIF from the culture media wild type cells began to adopt the flatter morphology of neurectoderm (Fig. 5A)(Ying et al., 2003). This was accompanied by downregulation of pluripotency-associated genes and the activation of a neural gene expression programme (Fig. 5B). *Mbd3Δ* ES cells are able to respond to the absence of self-renewal factors but have a very low probability of adopting a differentiated fate when induced to differentiate in N2B27 conditions (Kaji et al., 2006, Reynolds et al., 2012). Consistent with these findings, after 5 days in differentiation conditions *Mbd3Δ* ES cells showed some signs of having responded to differentiation conditions, but still retained pockets of morphologically undifferentiated cells (Fig. 5A, Middle). In contrast the completely NuRD-null *Mta123Δ* ES cells appeared to have all exited

328 the self-renewal programme and adopted a flat, monolayer morphology (Fig. 5A, Bottom).
329 The ability of *Mta123Δ* ES cells to undergo morphologically normal neuroectodermal
330 differentiation was rescued upon re-expression of either Mta1, Mta2 or Mta3 (Fig. EV5A).
331 Since the differentiation process may be considered as a combination of exit from self-
332 renewal (downregulation of pluripotency-associated genes) and acquisition of lineage
333 specific gene expression programs, we focussed on changes to these two classes of genes by
334 RT-qPCR over a differentiation time course. Pluripotency-associated genes such as *Esrrb*,
335 *Nanog* and *Pou5f1* were downregulated during differentiation of wild type, *Mbd3*Δ and
336 *Mta123Δ* ES cells (Fig. 5B). While this downregulation was not absolutely dependent on the
337 presence of functional NuRD, the magnitude and kinetics of the response varied in a gene-
338 dependent manner. Genes associated with acquisition of neural fate such as *Nestin*, *Pax6*,
339 *Ascl1* and *Cdh2* were activated in the presence of wild type NuRD, but this response was
340 reduced or absent in the NuRD mutant lines (Fig. 5B). In contrast, *Mta123Δ* ES cells
341 aberrantly expressed markers of inappropriate lineages, such as *Elf5* and *Gata3*
342 (trophectoderm—but not *Cdx2* or *Eomes*), *Ebf4*, *Nobox* and *Hcn4* (genes associated more
343 generally with later stages of differentiation or transmembrane transport). Misexpression of
344 these genes was also seen in *Mbd3*Δ cells, although not always to the same extent.
345 Exogenous expression of either Mta1, Mta2 or Mta3 was able to rescue the ability of
346 *Mta123Δ* ES cells to activate neural gene expression to different extents during
347 differentiation (Fig. EV5B), consistent with their ability to rescue morphological phenotypes.
348 The morphology of *Mta123*Δ cells induced to differentiate towards neuroectoderm
349 was quite different from that of *Mbd3*Δ cells (Fig. 5A), but there was no large-scale
350 difference in gene categories changing between the two mutant lines after 48h in
351 differentiation conditions (Fig. EV5C). In general *Mta123*Δ cells misexpressed more genes

352 from each category and to a greater extent than did *Mbd3*Δ cells (Fig EV5D), indicating that
353 their abnormal morphology in differentiation conditions is most likely due to widespread
354 gene misexpression generally, rather than activation of a specific differentiation pathway.
355 *Mta123Δ* ES cells induced to differentiate towards a mesoderm fate similarly failed to
356 appropriately activate differentiation markers, and exogenous expression of individual MTA
357 proteins was again able to rescue the defect to varying degrees (Fig. EV5E).
358 To assess the differentiation potential of *Mta123*Δ ES cells in an unbiassed manner,
359 we allowed control, *Mbd3*Δ and *Mta123*Δ ES cells to differentiate in embryoid bodies in
360 basal media supplemented with serum (Doetschman et al., 1985). After 12 days control cells
361 had formed colonies containing numerous morphologically distinguishable cell types,
362 silenced pluripotency-associated gene expression and activated expression of genes
363 associated with all three germ layers (Fig. 5C, D). In contrast both *Mbd3*Δ and *Mta123*Δ cells
364 showed less vigorous outgrowths consisting predominantly of a monolayer of flat cells (Fig.
365 5C). While both mutant lines were able to silence expression of pluripotency-associated
366 genes, both showed reduced activation of differentiation-associated genes, with *Mta123*Δ
367 cells generally showing more profound expression defects than *Mbd3*Δ cells (Fig. 5D).
368 Together these data show that while NuRD is not strictly required for silencing of
369 pluripotency genes, its activity is required for proper activation of lineage-appropriate genes
370 and repression of lineage-inappropriate genes during differentiation.
371 To better understand how NuRD-null cells respond when induced to differentiate we
372 compared their gene expression profiles to a transcription landscape made from single cells
373 taken from early mouse embryos (Fig. 5E; EV5F, G) (Mohammed et al., 2017). In the self-
374 renewing state, *Mbd3Δ* and *Mta123Δ* cells clustered near control (parental) lines near
375 embryonic day 3.5 (E3.5) and E4.5 inner cell mass cells, as is expected for naïve mouse ES

cells (Boroviak et al., 2014). After 48 hours of differentiation control ES lines clustered with E5.5 epiblast cells, indicating that the ES cell differentiation process occurred analogously to development *in vivo*. While *Mbd3⊿* ES cells appear to have exited the self-renewing state and to have taken the same differentiation trajectory as wild type cells (i.e. leftwards along PC1; Fig 5E, EV5G), they did not travel as far along this trajectory as wild type cells, instead occupying a space between the E4.5 and E5.5 epiblast states. *Mta123⊿* cells travelled even less far along PC1 than *Mbd3Δ* ES cells, and rather than maintain the appropriate differentiation trajectory they also travelled along PC2, occupying a space between E4.5 epiblast and E4.5 primitive endoderm. This further demonstrates that not only is NuRD important for cells to be able to adopt the appropriate gene expression programme for a given differentiation event, but it is also important for cells to maintain an appropriate differentiation trajectory.

If *Mta123Δ* ES cells were undergoing a specific trans-differentiation event towards trophectoderm during ES cell differentiation then this could become more pronounced if exposed to normal differentiation conditions in a chimaeric embryo. If, in contrast, they are simply unable to differentiate properly, they would not be expected to contribute to early embryos. To distinguish between these possibilities, we assessed the ability of *Mta123Δ* ES cells to differentiate in chimaeric embryos. Equal numbers of control or *Mta123Δ* ES cells expressing a fluorescent marker were aggregated with wild type morulae and allowed to develop for 48 hours. While wild type cells contributed to the ICM of host embryos with 100% efficiency, *Mta123Δ* ES cells showed significantly reduced contribution and increased levels of apoptosis (Fig. 6A-B, EV5H). Those *Mta123Δ* cells that did survive in blastocysts were predominantly, but not always found in the inner cell mass. These cells expressed Sox2 but not Cdx2, indicating that they had not undergone inappropriate differentiation towards

400 a trophectoderm fate (Fig. 6A, B). Nevertheless, no *Mta123*Δ cells could be found in E5.5
401 dpc embryos, indicating that they were unable to adapt to developmental cues and form
402 epiblast. This is most consistent with *Mta123Δ* cells being unable to enter a normal
403 differentiation path in an ICM environment. The ability to contribute to the ICMs of
404 chimaeric embryos was rescued by constitutive expression of Mta2 in *Mta123Δ* cells (Fig.
405 6B). We therefore propose that NuRD functions not only to establish the correct lineage
406 identity of cells during the differentiation process, but also prevents inappropriate gene
407 expression to maintain an appropriate differentiation trajectory (Fig. 6C).
408

Discussion

Here we provide a biochemical and genetic dissection of the core NuRD component MTA proteins in mouse ES cells. In contrast to what has been found in somatic cell types, MTA proteins are not mutually exclusive in ES cell NuRD complexes and all combinations of MTA homo- and hetero-dimers can exist within NuRD. Different MTA proteins exhibit subtle differences in chromatin localisation or biochemical interaction partners, but we find no evidence for protein-specific functions in self-renewing or differentiating mouse ES cells. ES cells completely devoid of MTA proteins are complete NuRD nulls and are viable but show inappropriate expression of differentiation-associated genes, are unable to maintain an appropriate differentiation trajectory and do not contribute to embryogenesis in chimaeric embryos.

Protein subunit diversity is often found in chromatin remodelling complexes which specifies functional diversity (Hargreaves & Crabtree, 2011, Hota & Bruneau, 2016, Morey et al., 2012). This is also the case for the NuRD complex, where alternate usage of Mbd2/3, Chd3/4/5, and Mta1/2/3 has been found to result in alternate functions for NuRD complexes (Feng & Zhang, 2001, Fujita et al., 2003, Nitarska et al., 2016). Yet our findings that Mta1 and Mta3-null mice are viable and fertile, and that we detect no major differences in the abilities of different MTA proteins to rescue the *Mta123Δ* ES cell phenotypes indicates that MTA proteins exhibit considerable functional redundancy during development. The MTA protein family expanded from a single copy to three orthologues near the base of the vertebrate clade. Triplication of this protein family may have allowed early vertebrates to diversify NuRD function in specific somatic tissues, conferring a selective advantage. Our finding that MTA proteins are functionally redundant in early

embryonic development indicates that the ancestral, essential function of the MTA proteins in early development has been retained by all three mammalian orthologues.

Different MTA proteins are capable of interacting with each other in ES cells, so how the mutual exclusivity reported in other cell types might be achieved is not clear. One possibility is that the variable inclusion in NuRD of zinc finger proteins, such as the Sall proteins in ES cells, could influence the MTA makeup of NuRD complexes. This class of variable NuRD interactors, which include Sall1/2/3/4, Zfp423 (Ebfaz), Zfpm1/2 (Fog1/2), and Bcl11b, interact with RBBP and/or MTA proteins via a short N-terminal motif (Hong et al., 2005, Lauberth & Rauchman, 2006, Lejon et al., 2011). Of this class of proteins Sall1 and Sall4 are the most highly expressed in ES cells, and Sall4 can associate with all three MTA proteins (Fig. 1C) (Miller et al., 2016). In contrast the Zfpm1 (Fog1) protein was shown to preferentially associate with Mta1 and Mta2, but not Mta3, in a somatic cell line (Hong et al., 2005). Hence it is possible that different proteins using this N-terminal motif to interact with NuRD in different cell types could act to skew the proportion of different MTA proteins included in the NuRD complex.

This genetic dissection has revealed that NuRD component proteins display a large degree of inter-dependence for protein stability, as has been reported for other multiprotein complexes (Roumeliotis et al., 2017). Mta1 and Mta2 both contain two distinct RBBP interaction domains, while Mta3 lacks the C-terminal most RBBP interaction domain (Millard et al., 2016). The Mta1 and Mta2 proteins show minimal loss of stability in the absence of Mbd3, but Mta3 requires an interaction with Mbd3 to be completely stable (Fig. EV4B) (Bornelov et al., 2018). It is possible that the additional interaction with an Rbbp protein confers stability to Mta1 and Mta2 in the absence of Mbd3. This would be consistent with our interpretation that Mta3 preferentially exists within an intact NuRD

456 complex (Figs 1, 2). Rbbp4/7 confer histone H3 binding to the NuRD complex, so Mta3-
457 containing NuRD may bind chromatin less tightly than Mta1- and/or Mta2-containing NuRD.
458 We have used pluripotent cells lacking Mbd3 extensively to show that NuRD plays
459 important roles in control of gene expression during early stages of exit from pluripotency in
460 vivo and in culture (Kaji et al., 2006, Kaji et al., 2007, Latos et al., 2012, Reynolds et al.,
461 2012). *Mbd3*-null ES cells contain Mbd2/NuRD and thus represent a NuRD hypomorph,
462 rather than a NuRD-null. ES cells lacking all three MTA proteins show no detectable NuRD
463 formation (Fig. 3) and therefore we propose represent a true NuRD null. *Mta123Δ* ES cells
464 are similar to *Mbd3Δ* ES cells in that both misexpress a range of genes in 2iL conditions and
465 both fail to properly differentiate. Yet the *Mta123Δ* ES cells misexpress more genes and
466 generally to a greater extent than do *Mbd3Δ* cells in both self-renewing and differentiation
467 conditions, and they show more pronounced differentiation defect. We propose that this
468 phenotype is a direct consequence of NuRD's function in suppressing transcriptional noise
469 during cell fate decisions. Moderate amounts of noise in the *Mbd3*Δ state allows cells to
470 maintain a differentiation trajectory, but interferes with their ability to achieve the
471 appropriate differentiation state (Fig. 6C). In the higher levels of noise characteristic of a
472 complete absence of NuRD cells are unable to appropriately follow differentiation cues and
473 maintain a coherent differentiation trajectory (Fig. 6C). We therefore propose that NuRD
474 encodes a highly conserved, redundant function to prevent transcriptional noise, as well as
475 a further, Mbd3-dependent function to fine tune transcriptional outputs during
476 differentiation.
477
478 **Materials and Methods**
479 **Mouse embryonic stem cells**

Mouse embryonic stem cells (ESCs) were grown on gelatin-coated plates in 2i/LIF conditions as described (Ying et al., 2008). All lines were routinely PCR genotyped and tested for mycoplasma. All ES cell lines used in this study are listed in Table EV1.

The following "Knockout First" alleles were obtained from EUCOMM as heterozygous ES cell lines (Illustrated in Fig. EV1A, B):

$Mta1^{tm1a(EUCOMM)Wtsi}$

https://www.mousephenotype.org/data/alleles/MGI:2150037/tm1a(EUCOMM)Wtsi

$Mta2^{tm1a(EUCOMM)Wtsi}$

https://www.mousephenotype.org/data/alleles/MGI:1346340/tm1a(EUCOMM)Wtsi

ES cells were used to derive mouse lines by blastocyst injection using standard methods. ESC derivation was performed by isolating ICMs and outgrowing in 2i/LIF media as described (Nichols et al., 2009).

The epitope tagged ESC lines Mta1-Avi-3×FLAG and Mta2-GFP were generated by traditional gene targeting, while the Mta3-Avi-3×FLAG line was generated using a CRISPR/Cas9 genome editing approach. *Mta3* contains three alternate stop codons in exons 14, 16 and 17, with exon 15 accessed by alternate usage of a splice donor site in exon 14. Visual inspection of our RNAseq data on a genome browser showed very little transcription beyond exon 14 in ES cells, indicating that the short isoform was predominant (Appendix Figure S2). We therefore targeted the stop codon present in exon 14 and simultaneously destroyed the alternate splice donor site using a synonymous mutation (GAA TGT to GAA TGC) to prevent splicing around our inserted construct in the targeted allele. The following sequences were targeted by guide RNAs in this experiment: 5'-GGCAAGGAGCGGAACGCGGA-3' and 5'-GGATGGCAAGGAGCGGAACG-3'. All MTA epitope tagged lines were made in an $Mbd3^{Flox/-}$ background (Kaji et al., 2006).

*Mta123*Δ ES cells were made by targeting *Mta2* in *Mta1*$^{\Delta/\Delta}$ *Mta3*$^{Flox/Flox}$ ES cells. Guide RNAs targeting exons 2 and 18 of *Mta2* were transfected along with a targeting construct designed to replace exons 2-18 with a puromycin expression cassette. The resulting ES cell lines had replaced exon 2-18 with the selection cassette on one allele, and had a deletion of the same exons on the other allele. The following sequences were targeted by guide RNAs: exon 2 (5'-TAGACGTAATCTGTAGGAGG-3' and 5'-AAATAGACGTAATCTGTAGG-3') and exon 18 (5'-AAATGCGCCGAGCGGCCCGA-3' and 5'-TCACCTGGAAATGCGCCGAG-3').

ES cells were induced to differentiate towards a neuroectoderm fate by removal of two inhibitors and LIF and culturing in N2B27 media as described (Ying et al., 2008). Mesendoderm differentiation was performed as follows: ES cells were plated at 10^4 cells/cm^2 in N2B27 on fibronectin-treated 6-well plates and cultured for 48 hours. Medium was then replaced with 10 ng/ml activin A and 3 µM CHIR99021 in N2B27 and cultured further. Embryoid bodies were allowed to form by allowing ~500 cells to aggregate in hanging drops of N2B27 media supplemented with 10% fetal calf serum. After 3 days embryoid bodies were allowed to attach and outgrow on gelatin-coated plates in N2B27 + serum. Experiments were carried out in quadruplicate.

Mice

All animal experiments were approved by the Animal Welfare and Ethical Review Body of the University of Cambridge and carried out under appropriate UK Home Office licenses. The Mta3 "Knockout First" allele was obtained from EUCOMM as a heterozygous mouse line (Illustrated in Fig. EV1C):

Mta3$^{tm3a(KOMP)Wtsi}$

https://www.mousephenotype.org/data/alleles/MGI:2151172/tm3a%2528KOMP%2529Wtsi?

Heterozygous knockout first mouse lines were crossed to a Flp-deletor strain kindly provided by Andrew Smith (University of Edinburgh) (Wallace et al., 2007) to generate conditional alleles. Mice harbouring conditional alleles were crossed to a Sox2-Cre transgenic line (Hayashi et al., 2002) to create null alleles. Mouse lines created in this study are listed in Table EV1.

Chimaeric embryos were made by morula aggregation with 8-10 ES cells per embryo as described (Hogan et al., 1994) and cultured for 24-48 hours prior to fixation and immunostaining. ES cells used stably expressed a PiggyBac Kusabira Orange transgene.

Nuclear Extracts and Immunoprecipitation

For extraction of nuclear proteins, ES cells were lysed in ice cold Buffer A (10mM HEPES pH7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.5mM DTT) supplemented with protease inhibitors and containing 0.6% NP-40. Nuclear pellets were washed further with Buffer A to reduce cytoplasmic protein contamination before extraction by vigorous shaking at 4° C for 1 hour in ice cold Buffer C (10mM HEPES pH7.9, 12.5% glycerol, 0.75mM $MgCl_2$, 10mM KCl, 0.5mM DTT) with protease inhibitors followed by centrifugation to remove insoluble fraction.

For immunoprecipitation of Flag-tagged proteins, anti-Flag antibody was incubated with Protein G-Sepharose beads (Sigma) for 1h at room temperature. Nuclear extract (200 µg) was diluted in IP-Buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% DOC) in the presence of protease inhibitors and Benzonase and incubated with antibody-bead conjugates at 4° C overnight. Beads were washed three times in ice cold IP Buffer and a further two times in IP Buffer containing 300mM NaCl.

Chd4 immunoprecipitations were carried out at 4° C overnight using antibody to endogenous Chd4 with Protein G sepharose in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 10% glycerol and washed as for the Flag IPs before SDSPAGE.

For immunoprecipitation of GFP-tagged proteins, Chromotek GFP-Trap agarose beads were used. Nuclear extract (200µg) was incubated with 25µl GFP-Trap beads in GFP IP buffer (10mM Tris pH 7.5, 150mM NaCl, 0.5mM EDTA) with Benzonase and protease inhibitors for 1 hour at 4° C. Beads were washed five times in IP buffer prior to Western blotting.

IPs were carried out in bulk and loaded on multiple SDS gels for Western blot analysis with various antibodies. Antibodies are listed in Table EV2.

Label-free pulldowns and label-free quantitation (LFQ) LC-MS/MS analysis

Label-free GFP pulldowns were performed in triplicate as previously described (Kloet et al., 2018). 2 mg of nuclear extract was incubated with 7.5 µl GFP-Trap beads (Chromotek) or 15 µl Flag-sepharose beads (Sigma) and 50 µg/mL ethidium bromide in Buffer C (300 mM NaCl, 20 mM HEPES/KOH, pH 7.9, 20% v/v glycerol, 2 mM MgCl2, 0.2 mM EDTA) with 0.1% NP-40, protease inhibitors, and 0.5 mM DTT in a total volume of 400 µl. After incubation, 6 washes were performed: 2 with Buffer C and 0.5% NP-40, 2 with PBS and 0.5% NP-40, and 2 with PBS. Affinity purified proteins were subject to on-bead trypsin digestion as previously described (Smits et al., 2013). In short, beads were resuspended in 50 µl elution buffer (2 M urea, 50 mM Tris pH 7.5, 10 mM DTT) and incubated for 20 min in a thermoshaker at 1400 rpm at room temperature. After addition of 50 mM iodoacetamide (IAA), beads were incubated for 10 min at 1400 rpm at room temperature in the dark. Proteins were then on-bead digested into tryptic peptides by addition of 0.25 µg trypsin (Promega) and subsequent

incubation for 2 h at 1400 rpm at room temperature. The supernatant was transferred to new tubes and further digested overnight at room temperature with an additional 0.1 µg of trypsin. Tryptic peptides were acidified and desalted using StageTips (Smits et al., 2013) prior to mass spectrometry analyses.

Chromatin immunoprecipitation (ChIP), sequencing, and analysis

Chromatin immunoprecipitations were carried out as previously outlined (Reynolds et al., 2012). For sequencing of ChIP DNA, samples from six (Mta1-FLAG and Mta2-GFP) and four (Mta3-FLAG) individual ChIP experiments were used. Antibodies used are listed in Table EV2. ChIP-seq libraries were prepared using the NEXTflex Rapid DNA-seq kit (Illumina) and sequenced at the CRUK Cambridge Institute Genomics Core facility (Cambridge, UK) on the Illumina platform. Low-quality reads and adaptor sequences were removed by *Trim Galore!* (version: 0.4.1, https://github.com/FelixKrueger/TrimGalore). The trimmed ChIP-seq reads were aligned to the mouse reference genome (GRCm38/mm10) using *bowtie* (version: 0.12.8; options: -m 1 -v 1) (Langmead, 2010). This was followed by peak calling using *macs2 callpeaks* against the corresponding inputs using default parameters and a Q-value of < 0.05. The R/Bioconductor *Diffbind* package was used to compare each combination of bound Mta proteins against all others using cut-off values of RPKM > 1.5, logFC > 1 and FDR < 0.05. The peaks were processed and visualized by *deepTools computeMatrix* and *plotHeatmap (Ramirez et al., 2014)*. Venn diagrams were plotted using the R/Bioconductor *ChIPseeker* package.

DNA Methylation data was obtained from (Shirane et al., 2016), low-quality reads and adaptor sequences were removed by *Trim Galore!* (version: 0.4.1), and reads were aligned to the mouse reference genome (GRCm38/mm10) using *Bismark* (version: 0.20.0)(Krueger &

599 Andrews, 2011). Alignments were de-duplicated and further processed using *Methpipe*.
600 Only CpGs with a coverage of at least 3 reads were considered. CpG methylation levels were
601 visualized by *deepTools* using a sliding window of 250 bp to give a smoothed track for the
602 heatmaps and profiles. Sequencing datasets are listed in Table EV3.
603
604 **Gene expression analyses**
605 Total RNA was purified using RNeasy Mini Kit (Qiagen) including on-column DNase
606 treatment. First-strand cDNA was synthesized using SuperScript IV reverse transcriptase
607 (Invitrogen) and random hexamers. Quantitative PCR (qPCR) reactions were performed
608 using TaqMan reagents (Applied Biosystems) on a QuantStudio Flex Real-Time PCR System
609 (Applied Biosystems) or a StepOne Real-Time PCR System (Applied Biosystems). Gene
610 expression was determined relative to housekeeping genes using the ΔCt method. TaqMan
611 assays, or PCR primers are listed in Table 1.
612 Table 1. List of TaqMan assays and primers used for gene expression analyses.

Gene	**TaqMan assay**
Ascl1	Mm03058063_m1
Atp5a1	Mm00431960_m1
Bmp4	Mm00432087_m1
Cdh2	Mm01162497_m1
Cdx2	Mm01212280_m1
Esrrb	Mm00442411_m1
Eomes	Mm01351985_m1
Elf5	Mm00468732_m1
Foxa2	Mm01976556_s1
Gapdh	Mm99999915_g1
Gata3	Mm00484683_m1
Gata6	Mm00802636_m1
Klf4	Mm00516104_m1
Nanog	Mm02019550_s1
Nestin	Mm00450205_m1
Nodal	Mm00443040_m1
Pax6	Mm00443081_m1

Pou5f1	Mm03053917_g1
Ppia	Mm02342430_g1
Sox1	Mm00486299_s1
Sox17	Mm00488363_m1
T	Mm00436877_m1
Tbp	Mm00446971_m1
Zfp42	Mm03053975_g1

Gene	**Forward primer (5' – 3')**	**Reverse primer (5' – 3')**
Ebf4	AACTGCGGGTCCTTATGTTC	TTTGTCTTTTCCGTCCCAGG
Hcn4	CCCGCCTCATTCGATACATT	AGCAGAAGCATCATGCCAAT
Nobox	CCTACGGAGAAGCTCTGCAA	TCTGGGGGAGAACAACCTTC
Sfrp2	AATGAGGACGACAACGACAT	ACGCCGTTCAGCTTGTAAAT

RNA-seq

Libraries for sequencing were prepared using the NEXTflex Rapid Directional RNA-seq kit (Illumina) or SMARTer® Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio) and sequenced on the Illumina platform as for ChIP-seq libraries. Low-quality reads and adaptor sequences were removed by *Trim Galore!* (version: 0.4.1), and the trimmed reads were aligned to mouse reference genome (GRCm38/mm10) using *Tophat* (version v2.1.0, (Kim et al., 2013)). Read counts per gene were obtained using *featureCounts* (version 1.5.0, (Liao et al., 2014) with annotations from Ensembl release 86 (Yates et al., 2016). Normalization, with and without spike-ins, and differential expression analyses were performed using the R/Bioconductor Deseq2 package (version 1.14.1, (Love et al., 2014)). The *R prcomp* function was used for principal component analysis (PCA). Gene Ontology (GO) analysis was performed using DAVID (Huang da et al., 2009b).

Statistical Analyses

No power calculations were performed to predetermine sample sizes. Statistical

630 analyses were performed using GraphPad Prism software. Images of cells or embryos are
631 representative of at least three independent experiments. Statistical methods used to
632 analyse high throughput sequencing data and proteomics data are described in the
633 respective sections of the Methods.
634
635 **Data Availability**
636 High throughput sequence datasets used in this manuscript are listed in Table EV3.
637 ChIP-seq data are available with the GEO accession number GSE122833. RNA-seq data are
638 available with the GEO accession number GSE122696. The mass spectrometry proteomics
639 data have been deposited to the ProteomeXchange Consortium via the PRIDE partner
640 repository with the dataset identifier PXD009855. All original western blot images are
641 available at Mendeley Data: DOI: 10.17632/d25bgfn8hd.1.
642
643 **Acknowledgments**
644 We thank Bill Mansfield, Peter Humphreys, Maike Paramor, Vicki Murray, and Sally
645 Lees for technical assistance and advice, and Alexander Brehm, Austin Smith, Ernest Laue
646 and members of the BDH lab for discussions and comments on the manuscript. Funding to
647 the BH and MV labs was provided through EU FP7 Integrated Project “4DCellFate” (277899).
648 The BH lab further benefitted from a Wellcome Trust Senior Fellowship (098021/Z/11/Z)
649 and from core funding to the Cambridge Stem Cell Institute from the Wellcome Trust and
650 Medical Research Council (097922/Z/11/Z and 203151/Z/16/Z). The Vermeulen lab is part of
651 the Oncode Institute, which is partly funded by the Dutch Cancer Society (KWF).
652
653 **Author Contributions**

TB and BH devised the study; TB, SK, SG, RF, JC, NR and BH generated the data; MB, MR and SD analysed high throughput sequencing data, SK and MV generated and analysed proteomics data, MK provided methodology and NR and BH wrote the manuscript with input from other authors.

Conflict of Interest Statement

The authors declare no conflicts of interest.

References

Allen HF, Wade PA, Kutateladze TG (2013) The NuRD architecture. Cell Mol Life Sci 70: 3513-24

Anastassiadis K, Fu J, Patsch C, Hu S, Weidlich S, Duerschke K, Buchholz F, Edenhofer F, Stewart AF (2009) Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. Dis Model Mech 2: 508-15

Bode D, Yu L, Tate P, Pardo M, Choudhary J (2016) Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells. Mol Cell Proteomics 15: 878-91

Bornelov S, Reynolds N, Xenophontos M, Gharbi S, Johnstone E, Floyd R, Ralser M, Signolet J, Loos R, Dietmann S, Bertone P, Hendrich B (2018) The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. Mol Cell 71: 56-72 e4

Boroviak T, Loos R, Bertone P, Smith A, Nichols J (2014) The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat Cell Biol 16: 516-28

Bowen NJ, Fujita N, Kajita M, Wade PA (2004) Mi-2/NuRD: multiple complexes for many purposes. Biochim Biophys Acta 1677: 52-7

Covington KR, Fuqua SA (2014) Role of MTA2 in human cancer. Cancer Metastasis Rev 33: 921-8

Denner DR, Rauchman M (2013) Mi-2/NuRD is required in renal progenitor cells during embryonic kidney development. Dev Biol 375: 105-16

Denslow SA, Wade PA (2007) The human Mi-2/NuRD complex and gene regulation. Oncogene 26: 5433-8

686 Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R (1985) The in vitro development
687 of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands
688 and myocardium. J Embryol Exp Morphol 87: 27-45

689 Ee LS, McCannell KN, Tang Y, Fernandes N, Hardy WR, Green MR, Chu F, Fazzio TG (2017) An
690 Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5.
691 Stem Cell Reports 8: 1488-1496

692 Feng Q, Zhang Y (2001) The MeCP1 complex represses transcription through preferential
693 binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev 15: 827-832.

694 Fujita N, Jaye DL, Geigerman C, Akyildiz A, Mooney MR, Boss JM, Wade PA (2004) MTA3 and
695 the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. Cell 119: 75-
696 86

697 Fujita N, Jaye DL, Kajita M, Geigerman C, Moreno CS, Wade PA (2003) MTA3, a Mi-2/NuRD
698 Complex Subunit, Regulates an Invasive Growth Pathway in Breast Cancer. Cell 113: 207-219

699 Gomez-Del Arco P, Perdiguero E, Yunes-Leites PS, Acin-Perez R, Zeini M, Garcia-Gomez A,
700 Sreenivasan K, Jimenez-Alcazar M, Segales J, Lopez-Maderuelo D, Ornes B, Jimenez-
701 Borreguero LJ, D'Amato G, Enshell-Seijffers D, Morgan B, Georgopoulos K, Islam AB, Braun T,
702 de la Pompa JL, Kim J et al. (2016) The Chromatin Remodeling Complex Chd4/NuRD Controls
703 Striated Muscle Identity and Metabolic Homeostasis. Cell Metab 23: 881-92

704 Gunther K, Rust M, Leers J, Boettger T, Scharfe M, Jarek M, Bartkuhn M, Renkawitz R (2013)
705 Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and
706 binding to exon sequences. Nucleic Acids Res 41: 3010-21

707 Hargreaves DC, Crabtree GR (2011) ATP-dependent chromatin remodeling: genetics,
708 genomics and mechanisms. Cell Res 21: 396-420

709 Hayashi S, Lewis P, Pevny L, McMahon AP (2002) Efficient gene modulation in mouse
710 epiblast using a Sox2Cre transgenic mouse strain. Gene expression patterns : GEP 2: 93-7

711 Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A (2001) Closely related proteins MBD2
712 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev 15: 710-
713 23.

714 Ho L, Crabtree GR (2010) Chromatin remodelling during development. Nature 463: 474-84

715 Hogan B, Beddington R, Constantini F, Lacy E (1994) Manipulating the Mouse Embryo. Cold
716 Spring Harbor Laboratory Press, Plainview, NY

717 Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA (2005) FOG-1
718 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1.
719 EMBO J 24: 2367-78

720 Hota SK, Bruneau BG (2016) ATP-dependent chromatin remodeling during mammalian
721 development. Development 143: 2882-97

722 Huang da W, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: paths
723 toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13

724 Huang da W, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large
725 gene lists using DAVID bioinformatics resources. Nature protocols 4: 44-57

Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA, Hendrich B (2006) The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol 8: 285-92

Kaji K, Nichols J, Hendrich B (2007) Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development 134: 1123-32

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14: R36

Kloet SL, Baymaz HI, Makowski M, Groenewold V, Jansen PW, Berendsen M, Niazi H, Kops GJ, Vermeulen M (2015) Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. FEBS J 282: 1774-85

Kloet SL, Karemaker ID, van Voorthuijsen L, Lindeboom RGH, Baltissen MP, Edupuganti RR, Poramba-Liyanage DW, Jansen P, Vermeulen M (2018) NuRD-interacting protein ZFP296 regulates genome-wide NuRD localization and differentiation of mouse embryonic stem cells. Nat Commun 9: 4588

Knock E, Pereira J, Lombard PD, Dimond A, Leaford D, Livesey FJ, Hendrich B (2015) The methyl binding domain 3/nucleosome remodelling and deacetylase complex regulates neural cell fate determination and terminal differentiation in the cerebral cortex. Neural Dev 10: 13

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27: 1571-2

Lai AY, Wade PA (2011) Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat Rev Cancer 11: 588-96

Langmead B (2010) Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics Chapter 11: Unit 11 7

Latos PA, Helliwell C, Mosaku O, Dudzinska DA, Stubbs B, Berdasco M, Esteller M, Hendrich B (2012) NuRD-dependent DNA methylation prevents ES cells from accessing a trophectoderm fate. Biology Open 1: 341-352

Lauberth SM, Rauchman M (2006) A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. J Biol Chem 281: 23922-31

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, Stunnenberg HG (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol 26: 843-51

Lejon S, Thong SY, Murthy A, AlQarni S, Murzina NV, Blobel GA, Laue ED, Mackay JP (2011) Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48.FOG-1 complex. J Biol Chem 286: 1196-203

Liang Z, Brown KE, Carroll T, Taylor B, Vidal IF, Hendrich B, Rueda D, Fisher AG, Merkenschlager M (2017) A high-resolution map of transcriptional repression. Elife 6

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30: 923-30

767 Link S, Spitzer RMM, Sana M, Torrado M, Volker-Albert MC, Keilhauer EC, Burgold T,
768 Punzeler S, Low JKK, Lindstrom I, Nist A, Regnard C, Stiewe T, Hendrich B, Imhof A, Mann M,
769 Mackay JP, Bartkuhn M, Hake SB (2018) PWWP2A binds distinct chromatin moieties and
770 interacts with an MTA1-specific core NuRD complex. Nat Commun 9: 4300

771 Loughran SJ, Comoglio F, Hamey FK, Giustacchini A, Errami Y, Earp E, Gottgens B, Jacobsen
772 SEW, Mead AJ, Hendrich B, Green AR (2017) Mbd3/NuRD controls lymphoid cell fate and
773 inhibits tumorigenesis by repressing a B cell transcriptional program. J Exp Med 214: 3085-
774 3104

775 Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for
776 RNA-seq data with DESeq2. Genome Biol 15: 550

777 Low JK, Webb SR, Silva AP, Saathoff H, Ryan DP, Torrado M, Brofelth M, Parker BL, Shepherd
778 NE, Mackay JP (2016) CHD4 Is a Peripheral Component of the Nucleosome Remodeling and
779 Deacetylase Complex. J Biol Chem 291: 15853-66

780 Lu X, Kovalev GI, Chang H, Kallin E, Knudsen G, Xia L, Mishra N, Ruiz P, Li E, Su L, Zhang Y
781 (2008) Inactivation of NuRD component Mta2 causes abnormal T cell activation and lupus-
782 like autoimmune disease in mice. J Biol Chem 283: 13825-33

783 Ma L, Yao Z, Deng W, Zhang D, Zhang H (2016) The Many Faces of MTA3 Protein in Normal
784 Development and Cancers. Curr Protein Pept Sci 17: 726-734

785 Manavathi B, Peng S, Rayala SK, Talukder AH, Wang MH, Wang RA, Balasenthil S, Agarwal N,
786 Frishman LJ, Kumar R (2007) Repression of Six3 by a corepressor regulates rhodopsin
787 expression. Proc Natl Acad Sci U S A 104: 13128-33

788 Matsuura T, Miyazaki S, Miyazaki T, Tashiro F, Miyazaki JI (2017) Zfp296 negatively regulates
789 H3K9 methylation in embryonic development as a component of heterochromatin. Sci Rep
790 7: 12462

791 Mazumdar A, Wang RA, Mishra SK, Adam L, Bagheri-Yarmand R, Mandal M, Vadlamudi RK,
792 Kumar R (2001) Transcriptional repression of oestrogen receptor by metastasis-associated
793 protein 1 corepressor. Nat Cell Biol 3: 30-7

794 Menafra R, Brinkman AB, Matarese F, Franci G, Bartels SJ, Nguyen L, Shimbo T, Wade PA,
795 Hubner NC, Stunnenberg HG (2014) Genome-wide binding of MBD2 reveals strong
796 preference for highly methylated loci. PLoS One 9: e99603

797 Millard CJ, Varma N, Saleh A, Morris K, Watson PJ, Bottrill AR, Fairall L, Smith CJ, Schwabe
798 JW (2016) The structure of the core NuRD repression complex provides insights into its
799 interaction with chromatin. Elife 5: e13941

800 Millard CJ, Watson PJ, Celardo I, Gordiyenko Y, Cowley SM, Robinson CV, Fairall L, Schwabe
801 JW (2013) Class I HDACs share a common mechanism of regulation by inositol phosphates.
802 Mol Cell 51: 57-67

803 Miller A, Ralser M, Kloet SL, Loos R, Nishinakamura R, Bertone P, Vermeulen M, Hendrich B
804 (2016) Sall4 controls differentiation of pluripotent cells independently of the Nucleosome
805 Remodelling and Deacetylation (NuRD) complex. Development 143: 3074-84

806 Mohammed H, Hernando-Herraez I, Savino A, Scialdone A, Macaulay I, Mulas C, Chandra T,
807 Voet T, Dean W, Nichols J, Marioni JC, Reik W (2017) Single-Cell Landscape of Transcriptional

Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep 20: 1215-1228

Mohd-Sarip A, Teeuwssen M, Bot AG, De Herdt MJ, Willems SM, Baatenburg de Jong RJ, Looijenga LHJ, Zatreanu D, Bezstarosti K, van Riet J, Oole E, van Ijcken WFJ, van de Werken HJG, Demmers JA, Fodde R, Verrijzer CP (2017) DOC1-Dependent Recruitment of NURD Reveals Antagonism with SWI/SNF during Epithelial-Mesenchymal Transition in Oral Cancer Cells. Cell Rep 20: 61-75

Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, Di Croce L (2012) Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell 10: 47-62

Narlikar GJ, Sundaramoorthy R, Owen-Hughes T (2013) Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell 154: 490-503

Nichols J, Silva J, Roode M, Smith A (2009) Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development 136: 3215-22

Nitarska J, Smith JG, Sherlock WT, Hillege MM, Nott A, Barshop WD, Vashisht AA, Wohlschlegel JA, Mitter R, Riccio A (2016) A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. Cell Rep 17: 1683-1698

O'Shaughnessy A, Hendrich B (2013) CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. Biochem Soc Trans 41: 777-82

O'Shaughnessy-Kirwan A, Signolet J, Costello I, Gharbi S, Hendrich B (2015) Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification. Development 142: 2586-97

Ostapcuk V, Mohn F, Carl SH, Basters A, Hess D, Iesmantavicius V, Lampersberger L, Flemr M, Pandey A, Thoma NH, Betschinger J, Buhler M (2018) Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. Nature 557: 739-743

Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T (2014) deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 42: W187-91

Reynolds N, Latos P, Hynes-Allen A, Loos R, Leaford D, O'Shaughnessy A, Mosaku O, Signolet J, Brennecke P, Kalkan T, Costello I, Humphreys P, Mansfield W, Nakagawa K, Strouboulis J, Behrens A, Bertone P, Hendrich B (2012) NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell 10: 583-94

Reynolds N, O'Shaughnessy A, Hendrich B (2013) Transcriptional repressors: multifaceted regulators of gene expression. Development 140: 505-12

Roumeliotis TI, Williams SP, Goncalves E, Alsinet C, Del Castillo Velasco-Herrera M, Aben N, Ghavidel FZ, Michaut M, Schubert M, Price S, Wright JC, Yu L, Yang M, Dienstmann R, Guinney J, Beltrao P, Brazma A, Pardo M, Stegle O, Adams DJ et al. (2017) Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. Cell Rep 20: 2201-2214

Sen N, Gui B, Kumar R (2014) Role of MTA1 in cancer progression and metastasis. Cancer Metastasis Rev 33: 879-89

Shimbo T, Du Y, Grimm SA, Dhasarathy A, Mav D, Shah RR, Shi H, Wade PA (2013) MBD3 localizes at promoters, gene bodies and enhancers of active genes. PLoS Genet 9: e1004028

Shirane K, Kurimoto K, Yabuta Y, Yamaji M, Satoh J, Ito S, Watanabe A, Hayashi K, Saitou M, Sasaki H (2016) Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. Dev Cell 39: 87-103

Si W, Huang W, Zheng Y, Yang Y, Liu X, Shan L, Zhou X, Wang Y, Su D, Gao J, Yan R, Han X, Li W, He L, Shi L, Xuan C, Liang J, Sun L, Wang Y, Shang Y (2015) Dysfunction of the Reciprocal Feedback Loop between GATA3- and ZEB2-Nucleated Repression Programs Contributes to Breast Cancer Metastasis. Cancer Cell 27: 822-36

Signolet J, Hendrich B (2015) The function of chromatin modifiers in lineage commitment and cell fate specification. FEBS J 282: 1692-702

Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A (2011) A conditional knockout resource for the genome-wide study of mouse gene function. Nature 474: 337-42

Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res 41: e28

Spruijt CG, Luijsterburg MS, Menafra R, Lindeboom RG, Jansen PW, Edupuganti RR, Baltissen MP, Wiegant WW, Voelker-Albert MC, Matarese F, Mensinga A, Poser I, Vos HR, Stunnenberg HG, van Attikum H, Vermeulen M (2016) ZMYND8 Co-localizes with NuRD on Target Genes and Regulates Poly(ADP-Ribose)-Dependent Recruitment of GATAD2A/NuRD to Sites of DNA Damage. Cell Rep 17: 783-798

Toh Y, Pencil SD, Nicolson GL (1994) A novel candidate metastasis-associated gene, mta1, differentially expressed in highly metastatic mammary adenocarcinoma cell lines. cDNA cloning, expression, and protein analyses. J Biol Chem 269: 22958-63

Wallace HA, Marques-Kranc F, Richardson M, Luna-Crespo F, Sharpe JA, Hughes J, Wood WG, Higgs DR, Smith AJ (2007) Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. Cell 128: 197-209

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Keenan S, Lavidas I, Martin FJ et al. (2016) Ensembl 2016. Nucleic Acids Res 44: D710-6

Ying QL, Stavridis M, Griffiths D, Li M, Smith A (2003) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol 21: 183-6

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A (2008) The ground state of embryonic stem cell self-renewal. Nature 453: 519-23

Zhang T, Wei G, Millard CJ, Fischer R, Konietzny R, Kessler BM, Schwabe JWR, Brockdorff N (2018) A variant NuRD complex containing PWWP2A/B excludes MBD2/3 to regulate transcription at active genes. Nat Commun 9: 3798

Zhang W, Aubert A, Gomez de Segura JM, Karuppasamy M, Basu S, Murthy AS, Diamante A, Drury TA, Balmer J, Cramard J, Watson AA, Lando D, Lee SF, Palayret M, Kloet SL, Smits AH, Deery MJ, Vermeulen M, Hendrich B, Klenerman D et al. (2016) The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. J Mol Biol 428: 2931-42

Figure 1. Biochemical characterisation of epitope-tagged MTA proteins

A. Schematic of MTA proteins with different protein domains indicated as coloured boxes, and the C-terminal epitope tags indicated. The model below shows the MTA- and HDAC-containing subunit joined to the CHD4-containing subunit through the MBD protein.

B. Tagged MTA proteins in heterozygously targeted cell lines were immunoprecipitated and subsequent western blots probed with antibodies indicated at right. Solid black triangles indicate the locations of untagged proteins and open triangles show the position of epitope-tagged proteins in the anti-Mta1, Mta2 or Mta3 panels; and a grey triangle shows the location of IgG bands in the anti-Mbd3 panel. 10% of the input was loaded into “Input” lanes. Molecular weights in kDa are shown at left. (Full western blot images are available in Mendeley Data.)

C. Proteins co-purifying with Mta1, Mta2 or Mta3 in IP-mass spectrometry experiments in wild type cells (top) or Mbd3-null cells (bottom). Proteins showing significant enrichment with the bait protein are located outside the dotted lines. For all panels the protein being immunoprecipitated is indicated in red, the other MTA proteins in purple, and other NuRD components in blue. Each IP/Mass spec experiment set comprised three independent immunoprecipitations from one nuclear extract preparation.

D. Relative enrichment of indicated proteins in MTA pulldowns from wild type (WT) or *Mbd3*-null (Mbd3KO) ES cells, normalised to 2x MTA proteins. NuRD components comprising the remodelling subunit are labelled in blue, those comprising the deacetylase subunit in red. As MTA proteins are the bait in these experiments, they may be isolated more efficiently than their interaction partners. Error bars represent standard deviation from three (Mta1 and Mta2) or six (Mta3) replicate pulldowns.

Figure. 2. Mta proteins show similar chromatin binding patterns

A. ChIP-seq peaks identified for Mta1 and Mta2 (left), Mta1 and Mta3 (middle) or Mta2 and Mta3 (right) are plotted by enrichment for each protein. Peaks called for both proteins are indicated in the overlap.
B. Overlap of peaks identified by ChIP-seq for each MTA protein in wild type ES cells. Total peak numbers are indicated below each protein name. Each ChIP-seq dataset was made from biological triplicates.
C. Comparison of Mta1 and Mta2 peaks with Chd4 peaks, as in Panel (B).
D. ChIP-seq enrichment for indicated proteins or histone modifications is plotted across different subsets of Mta-bound sites. Mta1/2/3 refers to peaks identified in all three ChIP-seq datasets, whereas “Mta1 Only”, “Mta2 Only” or “Mta3 Only” refer to peaks only called for that protein.
E. Average enrichment of density plots in (D) are plotted for different subsets of Mta ChIP-seq peaks.

Figure 3. Mta1/Mta2/Mta3 triple null ES cells represent a complete NuRD KO

A. Western blots of wild type (Control), double or triple-null ES cell nuclear extract was probed with antibodies indicated at right. LaminB1 acts as a loading control. Approximate sizes are indicated in kDa.

B. Phase-contrast images of wild type or *Mta123Δ* ES cells in 2i+LIF conditions. Scale bars represent 100μm.

C. Western blots of anti-Chd4 immunoprecipitation of nuclear extract from indicated cell lines (top) probed with antibodies indicated at right. For the Gatad2b and Mbd3 blots, arrowheads indicate target proteins, while an asterisk marks a non-specific IgG bands. 10% of the input was loaded into "Input" lanes. Approximate sizes are indicated in kDa.

D. Western blot of a time course of *Mta3* deletion in *Mta1ΔMta2ΔMta3*$^{Flox/Flox}$*: Cre-ER* (*Mta12Δ3*$^{F/F}$) or Control (*Mta12Δ3*$^{F/F}$ without Cre-ER) ES cells probed for indicated NuRD component proteins or RNA Polymerase II as a loading control (αRNAPII). The time course is indicated at the top as Days + tamoxifen.

E. Western blot showing rescue of *Mta123*Δ ES cells by ectopic expression of Mta1, Mta2 or Mta3 from a transgene (TG). Total RNA Polymerase II acts as a loading control (αRNAPII). For all western blots molecular weight is indicated at left in kDa.

F. Model of NuRD complex structure. Upon loss of Mbd3 most of the NuRD complex falls apart into the Chd4-containing remodeller subcomplex and the MTA-containing deacetylase subcomplex, but some intact Mbd2-NuRD still remains. Upon loss of all three MTA proteins no intact NuRD remains, both Mbd3 and Gatad2a/b become unstable and neither of the intact subcomplexes remain.

Figure 4. Mta proteins act redundantly to control gene expression

A. Comparison of gene expression in *Mta123*Δ ES cells and wild type (WT) ES cells. Each circle represents a gene: red indicates spike-in controls, blue indicates genes that are not differentially expressed to a significant degree, and green indicates differentially expressed genes (2404 increased, 1293 decreased) defined with an adjusted p-value < 0.05 and a log2 fold-change > 1. N=3 for each genotype.

B. Fold change in gene expression is plotted for different subsets of genes. All genes are plotted in black, while subsets of genes located nearest to ChIP-seq peaks for the indicated proteins +Chd4 and which show significant changes in expression compared to wild type cells are plotted in dashed coloured lines as indicated. The number of genes in each of the Mta categories are: all genes (32271), all differentially expressed (n=3701), Mta1+2+3 (n=1738), Mta1+2 (n=1460), Mta1 (n=1020) and Mta2 (n=924).

C. Principal component analysis of RNAseq data from ES cell lines of indicated genotypes in either self-renewing conditions (2i) or after 48 hours in the absence of two inhibitors and LIF (Diff48). Each point represents a biological replicate.

D. GO term enrichment for genes differentially expressed in *Mta123*Δ ES cells (left) or *Mbd3Δ* ES cells (right) compared to wild type cells in 2iL conditions. For each comparison the top five most significant gene ontology terms are plotted by $\log_{10}$ of the Benjamini-adjusted p-value. The significant GO terms and p-values were calculated using David v.6.8

E. Genes associated with indicated GO terms plotted by fold change in expression in *Mta123Δ* ES cells (x-axis) or *Mbd3Δ* ES cells (y-axis). Genes are coloured if they are differentially expressed (log2 fold-change > 1 and adjusted p-value < 0.05) in either comparison as indicated. The dotted lines show the fold-change cut-off of 2. GO-

terms were identified using David v.6.8 (Huang da et al., 2009a) using a Benjamini score with a cutoff of 0.05.

F. Genes associated with core NuRD peaks, with an MTA-only peak, with a Chd4 peak and an MTA-peak but not an Mbd3 peak (Chd4+Mta-Mbd3 genes), or with an MTA-peak but not a Chd4 peak (MTA-only genes) were divided into quartiles based upon expression levels in wild type cells. Change in gene expression is plotted on the x-axis and calculated significance of the expression change on the y-axis. Those genes significantly misexpressed (|log2FC|>1; p≤0.05) are indicated in red, genes showing no significant change are plotted in black. The percentage of genes within each quartile which is significantly misexpressed is indicated at the top of each plot.

Figure 5. NuRD activity maintains an appropriate ES cell differentiation trajectory

A. Phase contrast pictures of ES cells of indicated genotypes after 5 days in neuroectoderm differentiation conditions (N2B27). White arrowheads in the *Mbd3*Δ panel indicate pockets of ES-like cells. Scale bars represent 100 µm.

B. Expression of indicated genes in the neuroectoderm differentiation time course was measured by RT-qPCR in Control (green) *Mta123Δ* (magenta) or *Mbd3*Δ (blue) ES cells. N ≥ 3 biological replicates, error bars indicate SEM.

C. Phase contrast images of embryoid body outgrowths of indicated genotypes after 12 days in differentiation conditions (N2B27 + 10% serum). Scale bars represent 100µm.

D. Expression of indicated genes in embryoid body cultures over time was measured by RT-qPCR as in Panel (B).

E. Comparison of expression data for wild type, *Mbd3Δ* or *Mta123Δ* ES cells in self-renewing (2i) conditions or after 48 hours in differentiation conditions (48h N2B27)

with mouse embryonic single cell RNA-seq data from (Mohammed et al., 2017). Larger shading encloses biological replicates, smaller circles represent individual cells. PC4 vs PC1 is plotted in Fig. EV5F and a loading plot in EV5G.

Figure 6. NuRD-null ES cells cannot contribute to normal embryonic development

A. Composite images of representative chimaeric embryos made with control (*Mta1*$^{Flox/\Delta}$*Mta2*$^{+/+}$*Mta3*$^{Flox/Flox}$) or *Mta123Δ* ES cells. ES-derived cells express the Kusabira Orange fluorescent marker. Sox2 indicates epiblast cells and Cdx2 is expressed in trophectoderm cells. Arrows indicate examples of K-Orange expressing cells in the mutant embryos. Scale bars = 20μm.

B. (Top) Number of K-Orange expressing cells observed in chimaeric embryos obtained using control ES cells, *Mta123Δ* ES cells, or *Mta123Δ* ES cells in which Mta2 was reintroduced on a constitutively expressed transgene (*Mta123Δ+Mta2*^{TG}). P-values calculated using a two-tailed t-test. (Middle) Mean number of K-Orange cells per embryo separated by Sox2 expression. P-values calculated using a two-tailed t-test. (Bottom) Number of K-Orange and Caspase-3 positive cells per embryo. P-values calculated using a one-tailed t-test: *P < 0.05, ****P < 0.0001, “ns” = not significant.

C. Model of NuRD function during differentiation of pluripotent cells. NuRD facilitates lineage commitment of ES cells after exit from pluripotency (red arrows), allowing cells to form differentiated cell types (Top). In the absence of Mbd3, residual NuRD activity ensures cells retain the appropriate differentiation trajectory, but the cells are unable to reach a differentiated cell fate (dotted red arrows; Middle). In the absence of all three MTA proteins there is no residual NuRD activity and ES cells are

unable to either achieve appropriate lineage commitment, or to maintain the proper differentiation trajectory (dotted arrows, Bottom).

Expanded View Figures

Figure EV1. Expression of MTA genes during early mouse development

RNAseq data from (Boroviak et al., 2014) plotted at indicated days of mouse development for each of the MTA genes. All data points are shown, with horizontal bars indicating the mean.

Figure EV2. Chromatin features of MTA-bound peaks

A. Comparison of ChIP-seq peaks for Mta3 and Chd4. The number of Mta3-only, Chd4-only, or Mta3+Chd4 peaks are indicated.
B. Features of peaks containing two of the three MTA proteins, as in Figure 2D.
C. Distribution of ChIP-seq peaks identified as shared amongst all three MTA proteins (MTA123) or as unique for indicated MTA proteins is plotted relative to Refseq transcription start sites (TSS) on the x-axis. Note distance is on a $\log_{10}$ scale.
D. Average enrichment of indicated features is shown for peaks for each Mta protein which do or do not also contain Chd4. The number of peaks in each set is indicated as n.
E. Average enrichment of DNA methylation across MTA peaks with or without Chd4, as in Panel (D).
F. Significance of GO-term enrichment for genes associated with peaks for different combinations of Mta proteins with or without Chd4.

Figure EV3. MTA reporter and knockout alleles

A. Schematic of the *Mta1* "Knockout First" reporter allele *Mta1*$^{tm1a(EUCOMM)Wtsi}$ (Top). Exons are depicted as boxes, with normal coding exons as filled boxes. Exons around the insertion site are numbered. Coding exons not able to be translated in the depicted allele are shaded in light blue. The targeting resulted in an FRT-flanked LacZ-Neo fusion protein being expressed from the endogenous *Mta1* promoter and preventing transcription of most *Mta1* exons. Recombination between FRT sites is achieved by expression of FLP recombinase (Middle), removing the LacZ-Neo cassette and restoration of Mta1 coding potential. Subsequent recombination between LoxP sites by Cre recombinase (Bottom) results in loss of Exon 2 and subsequent exons are out of frame.

B. Schematic of the *Mta2*$^{tm1a(EUCOMM)Wtsi}$ allele as in panel (A). In this allele the neo gene is expressed from a human ß-actin promoter and expression of Cre results in excision of exons 4-13.

C. Schematic of the *Mta3*$^{tm3a(KOMP)Wtsi}$ allele as in panel (A).

D. Western blot of nuclear extracts made from each single mutant probed with indicated antibodies. Anti-Rbp1 NTD (RNA polymerase II) acts as a loading control.

E. Phase-contrast images of cell lines of indicated genotype in self-renewing conditions. Scale bars indicate 100 µm.

F. Comparison of gene expression in indicated MTA single mutant ES cells and wild type (WT) ES cells. Each circle represents a gene: red indicates genes that are not differentially expressed to a significant degree, and green indicates differentially expressed genes. The number of up- and down-regulated genes defined with an

adjusted p-value < 0.05 and a log2 fold-change > 1 is indicated in each plot. N=3 for each genotype.

G. Overlap of misexpressed genes in indicated single mutant ES cells. The number of genes in each overlap category are indicated.

H. Gene expression data for each single MTA mutant ES line, the *Mbd3*Δ ES line and the *Mta123*Δ ES line and their control parental lines (WT) in self-renewing (2i) conditions are plotted relative to that of mouse embryonic single cell RNA-seq data from (Mohammed et al., 2017) (smaller grey circles).

Figure EV4. Control of gene expression by the MTA proteins

A. Western blot of nuclear extract from cell lines indicated at the bottom in 2iL conditions probed with antibodies listed at right. Arrows indicate the locations of the two Mbd2 isoforms in the anti-Mbd2 panel, and of the three Mbd3 isoforms in the anti-Mbd3 panel. The grey triangle indicates the location of Mbd3-3xFLAG, which was present in the *Mbd2*Δ line used. Anti-Rbp1 NTD (RNA polymerase II) acts as a loading control. Approximate sizes are indicated at left in kDa.

B. Western blot of indicated proteins in wild type and *Mbd3Δ* ES cells in 2iL conditions. Anti-Rbp1 NTD (RNA polymerase II subunit) acts as a loading control.

C. Comparison of gene expression changes in wild type (*Mta123WT* = control ES cell line derived at the same time as the parent of the *Mta123Δ* ES cell line), *Mbd3Δ* and *Mta123Δ* ES cells in self-renewing (2iL) conditions. The top 100 genes contributing to PC2 from Figure 5E are shown.

D. Common ChIP-seq peaks for Chd4 and all three MTA proteins (Core NuRD peaks) were assigned to the nearest genes. These genes were then divided into four equal

groups based upon expression levels in wild type cells, and their expression in *Mbd3*Δ ES cells (top panels) or *Mta123*Δ ES cells (bottom panels) is plotted relative to that in wild type ES cells. Change in gene expression is plotted on the x-axis and calculated significance of the expression change on the y-axis. Those genes significantly misexpressed (|log2FC|>1; p≤0.05) are indicated in red, genes showing no significant change are plotted in black. The percentage of misexpressed genes located within each quartile is indicated above each graph.

Figure EV5. Failure of differentiation in *Mta123*Δ ES cells

A. Phase contrast images of ES cells of indicated genotypes induced to differentiate for 5 days in N2B27. Scale bars represent 100 μm.
B. Comparison of gene expression in indicated ES cell lines in undifferentiated conditions or after 5 days in neuroectoderm differentiation conditions (N2B27). RT-qPCR was carried out in triplicate at each time point for a minimum of three biological replicates. Error bars indicate SEM.
C. GO term enrichment for genes differentially expressed in *Mbd3Δ* ES cells (left) or *Mta123*Δ ES cells (right) after 48h of differentiation in N2B27. For each comparison the top five most significant gene ontology terms are plotted by $\log_{10}$ of the Benjamini-adjusted p-value. Significant GO terms and p-values were calculated using David v.6.8 (Huang da et al., 2009a).
D. Expression of genes associated with indicated GO terms is plotted by fold change in expression in *Mta123Δ* ES cells (x-axis) or *Mbd3Δ* ES cells (y-axis) after 48h of differentiation in N2B27. Genes are coloured if they are differentially expressed (log2 fold-change > 1 and padj value < 0.05) in either comparison as indicated. The dotted

lines show the fold-change cut-off of 2. GO-terms were identified using David v.6.8 using a Benjamini score with a cutoff of 0.05.

E. Comparison of gene expression in indicated ES cell lines induced to differentiate towards mesoderm. qPCR was carried out in triplicate at each time point for a minimum of three biological replicates. Error bars indicate SEM.

F. Same as Figure 5E, plotting PC4 vs PC1.

G. Loading plot for Figure 5E. A selection of key genes is highlighted in different colours based upon a published classification of early embryonic gene expression (Boroviak et al., 2014).

H. Composite images of representative chimaeric embryos as in Figure 6A stained with an anti-activated Caspase 3 antibody. Arrows indicate an example of an orange, apoptotic cell. Scale bars = 20μm.

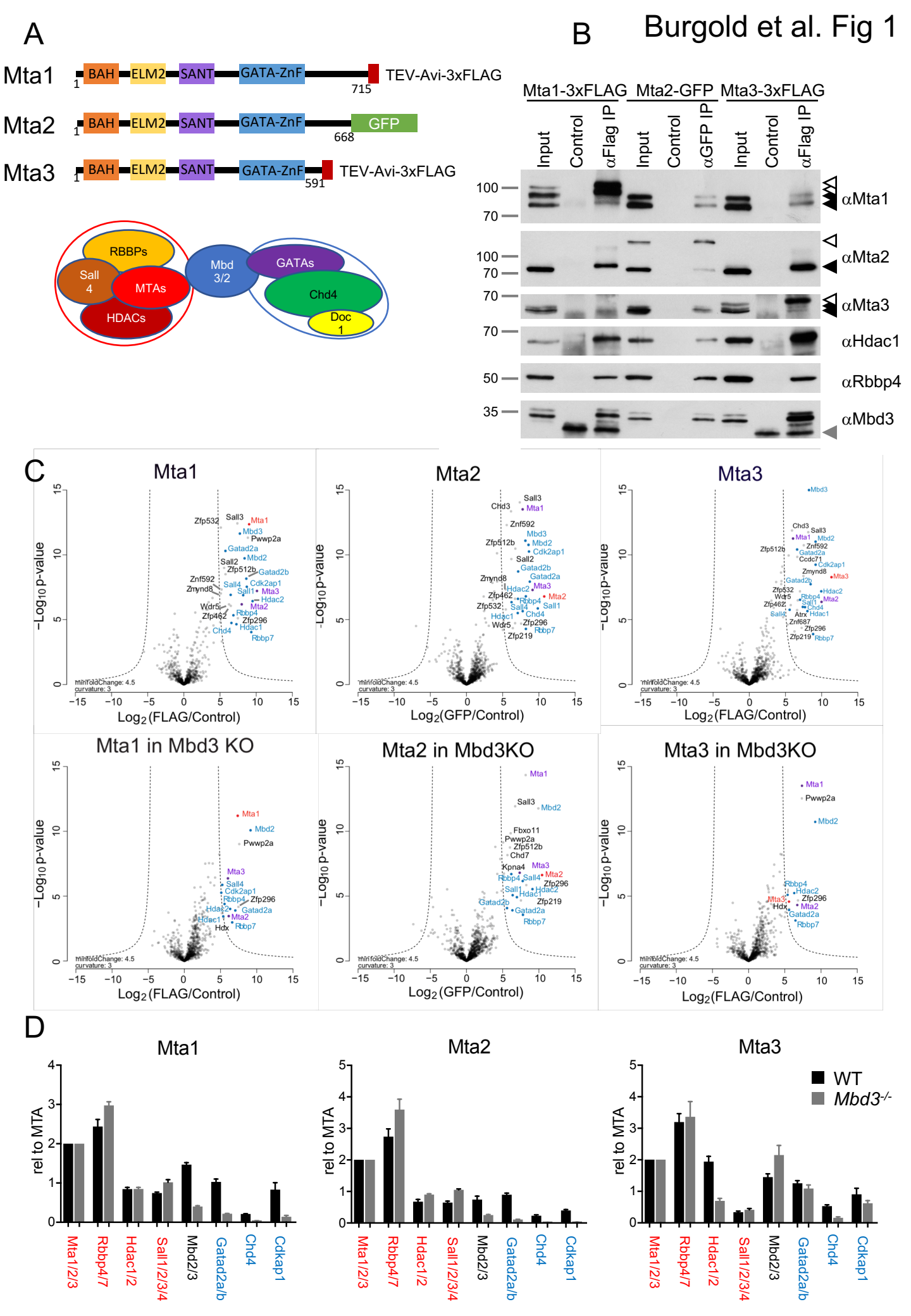
Table EV1. ES cell lines and mice

Table EV2. Antibodies

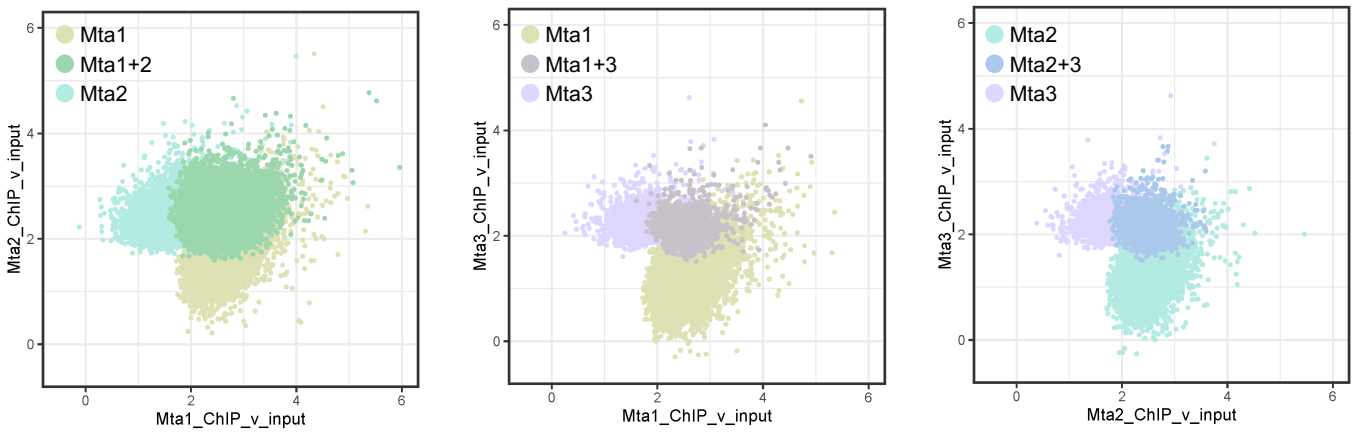
Table EV3. Sequencing Datasets

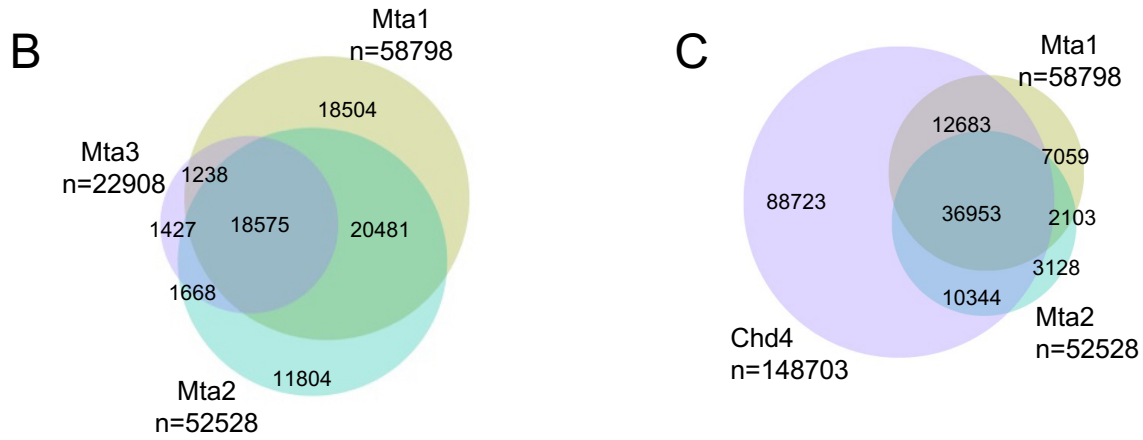
Appendix Figure 1. High correlation amongst replicates of ChIP-seq data

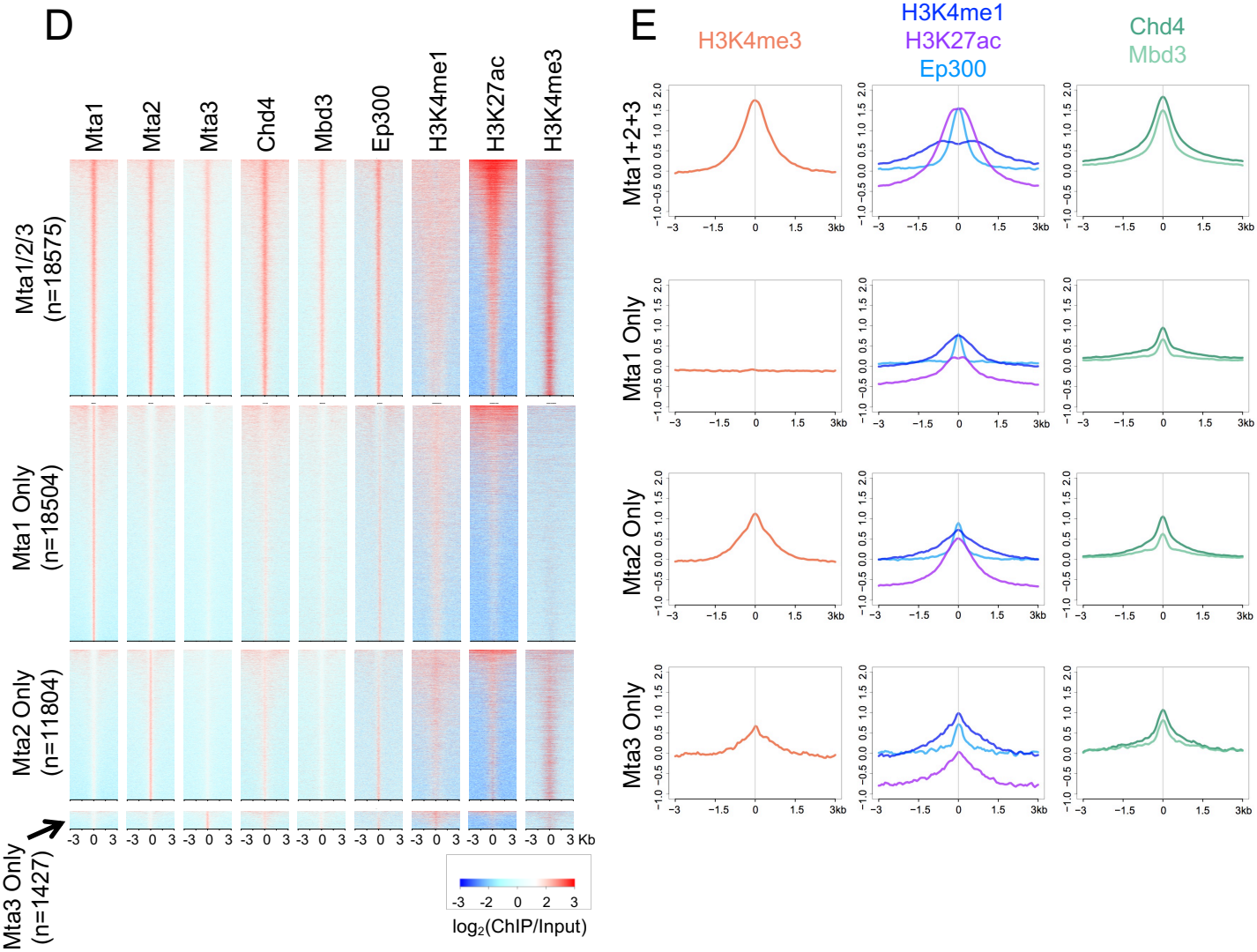
Appendix Figure 2. *Mta3* stop codon selection in ES cells

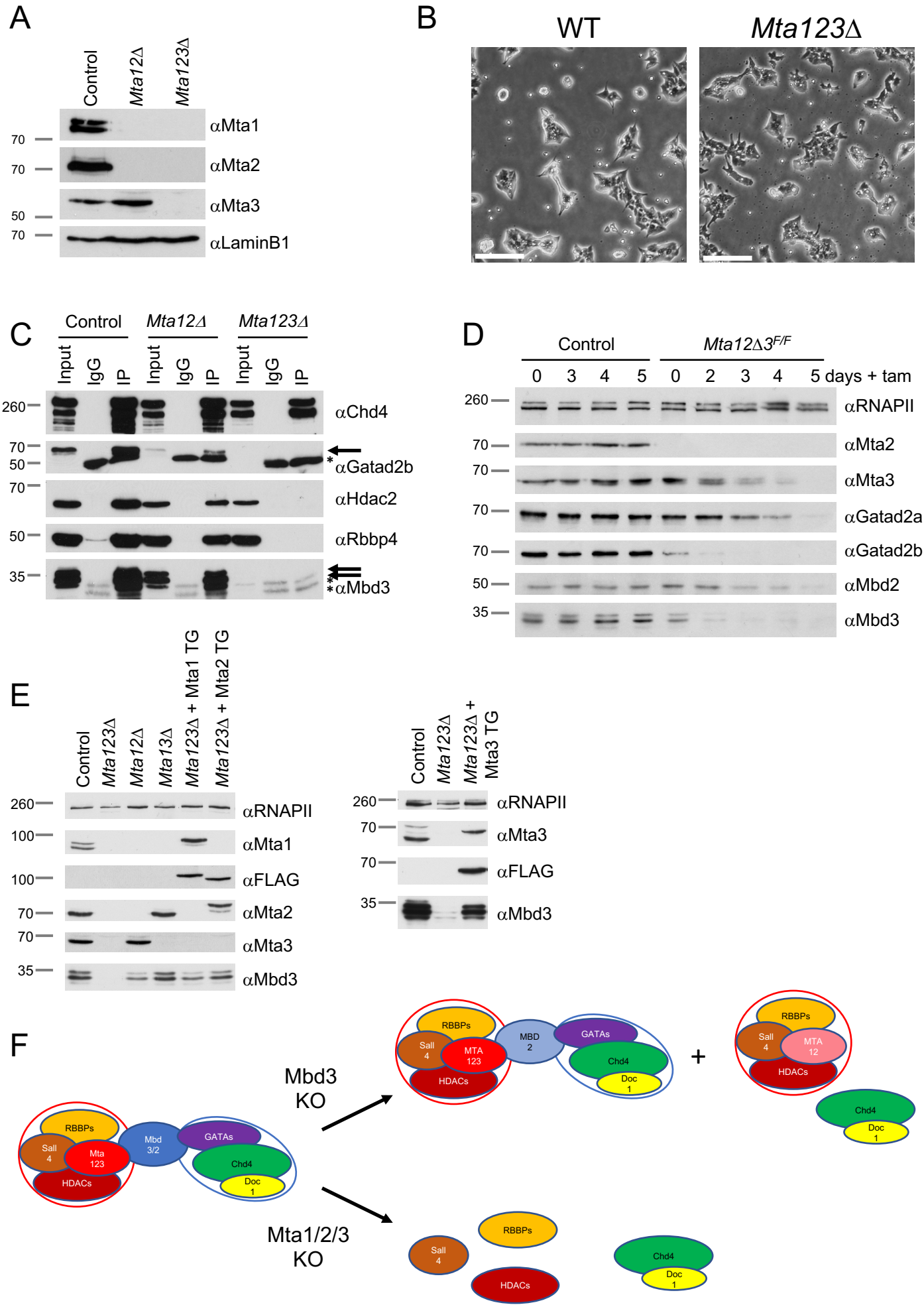

A
Mta1
BAH
ELM2
SANT
GATA-ZnF
TEV-Avi-3xFLAG
1
715
Mta2
GFP
668
Mta3
591
RBBPs
Sall 4
MTAs
HDACs
Mbd 3/2
GATAs
Chd4
Doc 1
B
Mta1-3xFLAG
Mta2-GFP
Mta3-3xFLAG
Input
Control
αFlag IP
αGFP IP
αMta1
αMta2
αMta3
αHdac1
αRbbp4
αMbd3
C
Mta1
Mta2
Mta3
Mta1 in Mbd3 KO
Mta2 in Mbd3KO
Mta3 in Mbd3KO
−Log10 p-value
Log2(FLAG/Control)
Log2(GFP/Control)
D
rel to MTA
WT
Mbd3-/-
Mta1/2/3
Rbbp4/7
Hdac1/2
Sall1/2/3/4
Mbd2/3
Gatad2a/b
Chd4
Cdkap1

Burgold et al. Fig 2

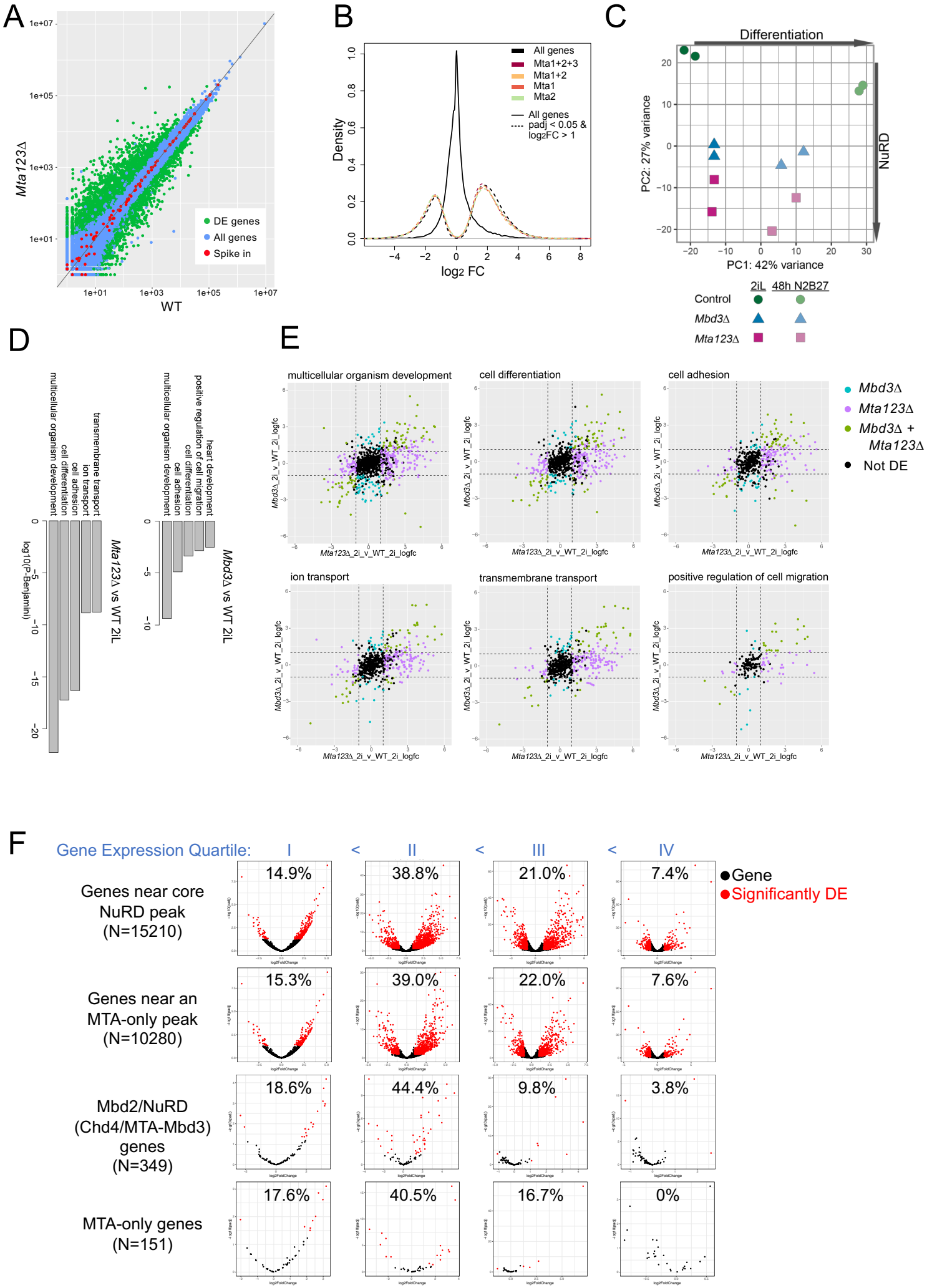




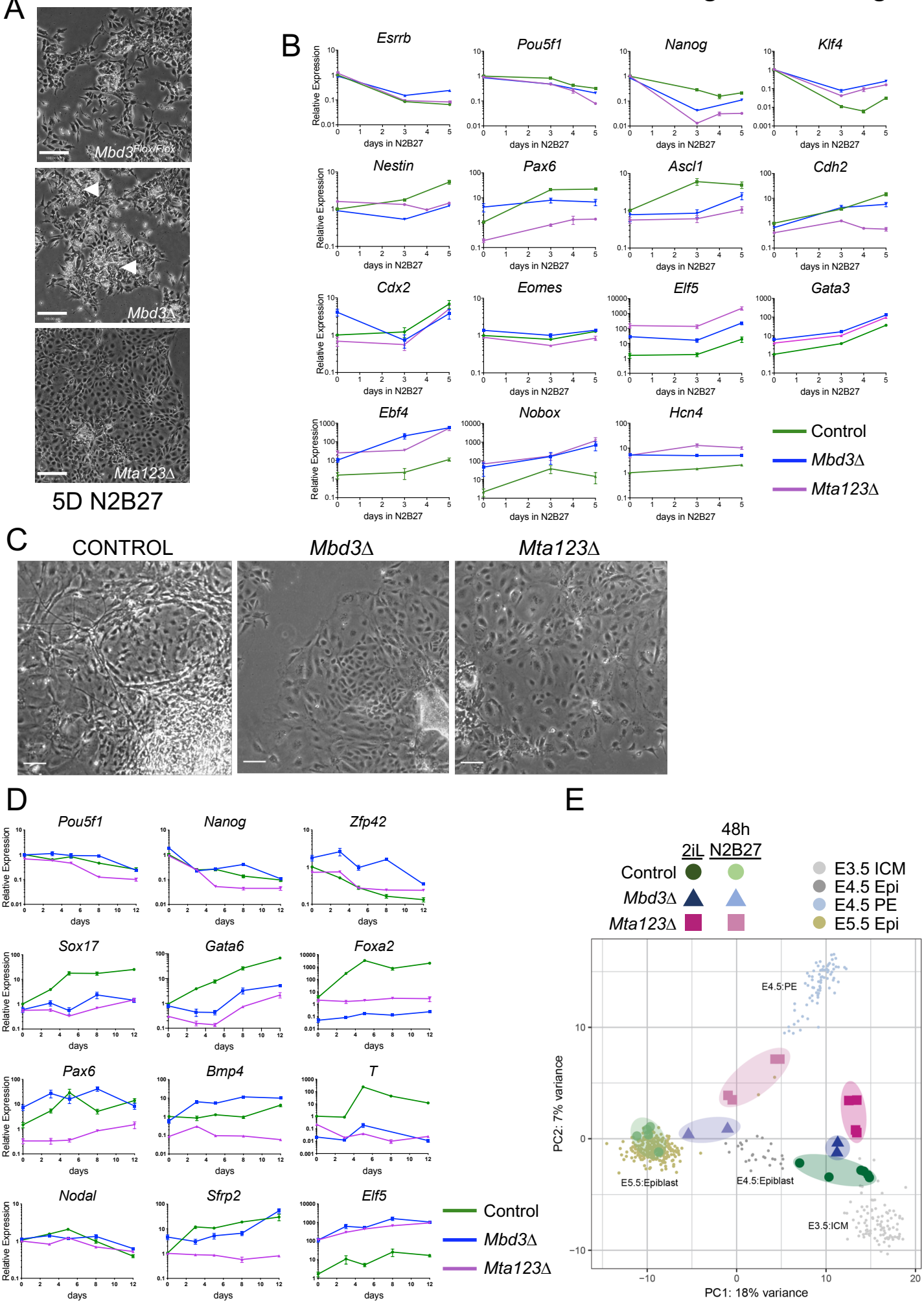



A
Control
Mta12Δ
Mta123Δ
70
70
50
70
αMta1
αMta2
αMta3
αLaminB1
B
WT
Mta123Δ
C
Control
Mta12Δ
Mta123Δ
Input
IgG
IP
260
70
50
70
50
35
αChd4
αGatad2b
αHdac2
αRbbp4
αMbd3
D
Control
Mta12Δ3F/F
0 3 4 5 0 2 3 4 5 days + tam
260
70
70
70
70
50
35
αRNAPII
αMta2
αMta3
αGatad2a
αGatad2b
αMbd2
αMbd3
E
Control
Mta123Δ
Mta12Δ
Mta13Δ
Mta123Δ + Mta1 TG
Mta123Δ + Mta2 TG
260
100
100
70
70
35
αRNAPII
αMta1
αFLAG
αMta2
αMta3
αMbd3
Control
Mta123Δ
Mta123Δ + Mta3 TG
260
70
70
35
αRNAPII
αMta3
αFLAG
αMbd3
F
Mbd3 KO
Mta1/2/3 KO
RBBPs
Sall 4
Mta 123
HDACs
Mbd 3/2
GATAs
Chd4
Doc 1
MBD 2
MTA 12
+

Burgold et al. Fig 4



Burgold et al. Fig 5



Burgold et al. Fig 6